ESTIMATING DEMAND WITH ATTITUDE: HOW OPINIONS AND FEELINGS AFFECT CONSUMER CHOICE

Gregory M. McAtee

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Economics.

Chapel Hill
2013

Approved by:

Dr. Donna Gilleskie

Dr. Saraswata Chaudhuri

Dr. David Guilkey

Dr. Helen Tauchen

Dr. William Putsis

Abstract

GREGORY M. MCATEE: ESTIMATING DEMAND WITH ATTITUDE: HOW OPINIONS AND FEELINGS AFFECT CONSUMER CHOICE.
(Under the direction of Dr. Donna Gilleskie.)

A consumer's mood, opinions, and general disposition can influence her demand for a product. While an empirical specification that models the correlation between attitudes and purchasing decisions will allow for a more accurate prediction of demand, there is disagreement among researchers over the role of attitudinal responses in estimation. This paper examines the relationship between the consumer's responses to questions about attitudes and demand for goods in order to demonstrate the most appropriate use of attitudinal data in empirical work. A theoretical model of consumer behavior identifies several latent factors that simultaneously influence the individual's purchasing decisions and responses to attitudinal questions. This correlation, driven by unobserved heterogeneity such as the individual's mood, opinions, discount factor, expectations over future outcomes, and unreported experiences, causes a simultaneity bias in specifications that include attitudinal responses as explanatory variables. Instead, I jointly estimate purchases and attitudinal responses using a random effects model to accurately capture the relationship between both observed outcomes. An econometric proof, a Monte Carlo experiment, and a data application show that, compared to commonly-used specifications, this jointly-estimated model improves the accuracy and efficiency of the estimated response parameters of the covariates that explain consumer demand.

Dedication

I dedicate this work to Angela L. DeSantis, Kathleen L. McAtee, Loretta M. Thompson, and my parents, Gregory B. and Angela D. McAtee, whose encouragement and support guided my pursuit of this advanced degree.

Acknowledgments

I am very grateful for the help and guidance I have received over the last few years. First and foremost, I would like to thank my dissertation advisor, Dr. Donna Gilleskie, for her expert advice and direction for this research. I am indebted to my dissertation committee members, Dr. David Guilkey, Dr. Helen Tauchen, Dr. Saraswata Chaudhuri, and Dr. William Putsis, for sharing their time with me to discuss the economic theory that supports my work. I am thankful for the thoughtful comments I received from Dr. Brian McManus and all participants of the UNC Applied Microeconomics Workshop; our insightful conversations contributed to my overall understanding of the ideas presented in this paper. I also thank Phil Khoury of the Cincinnati Children's Hospital Medical Center for his help obtaining the data.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

For any empirical application, the researcher's ability to quantify the determinants of an individual's behavior is limited by the contents of the data, which generally consist of responses to objective questions. While this information describes the individual's *observed* sociodemographic and environmental characteristics, we might find ourselves wondering to what extent her demand for goods and services is influenced by *intangible* personal characteristics: Did the individual's mood affect her decision? How forward looking is she? How does she form expectations over future outcomes that result from this decision? Typically, when these concerns enter our discussions, we acknowledge that these unobserved factors affect the outcome, regret that these data are not available, relegate their effects to the error component, and proceed with an assumption about the error term. But is there a better way to account for this unobserved heterogeneity?

It is becoming more common for surveys to include attitudinal questions that correspond to specific psychological traits. The participant's subjective responses can give the researcher an idea of the individual's mental state,[1] locus of control,[2] and opinions/beliefs,[3] to name a few. However, it is not clear how these subjective reports of one's well being might enter a standard model of economic decision making. In this paper, I construct a theoretical model to examine

---

[1]For example, measures of depression, stress, and anxiety are collected in the National Longitudinal Study of Youth and Add Health studies.

[2]For example, the Panel Study of Income Dynamics and the National Longitudinal Survey datasets record descriptions of participants' perceptions of the extent to which they can control events that affect them.

[3]For example, marketing surveys ask consumers about their attitudes towards a product or a product's characteristics.

the connection between attitudinal responses (r) and purchasing decisions (y). Ultimately, I conclude that the same exogenous characteristics and unobserved factors simultaneously affect both observed outcomes, and I derive a set of equations that, when jointly estimated, describes the significant determinants of the demand for good y and the production of attitudinal response r. The econometric specification accounts for observed and unobserved sources of correlation by including the same covariates in each equation and estimating the distribution of common unobserved factors across equations.

I use econometric theory and Monte Carlo simulations to explore the benefits of my estimation strategy. Relative to other specifications introduced later in the paper, the standard errors for all response parameters are more efficient when the outcome y and the attitudinal response r are jointly estimated. Furthermore, I show that the efficiency gain grows as the correlation among latent factors strengthens. To support this theory, I conduct a Monte Carlo experiment that measures the efficiency gains under several different data and error generating processes. Data from the National Heart, Lung, and Blood Institute Growth and Health Study (NGHS), a longitudinal survey of individuals that features a rich set of attitudinal questions, are used in an empirical application of the econometric and behavioral theory described in this paper. Because the survey provides an extensive history of each participant's smoking habits, the data application examines adolescent cigarette use. In this data application, the jointly-estimated model generates more accurate predictions of smoking behavior compared to some commonly-used specifications.

Chapter 2

Background and Contribution

This research draws from three topics in the literature: the use of attitudinal data in estimation, the role of unobserved heterogeneity, and the prediction of adolescent drug use. Having provided a brief literature review on each topic, I describe how previous research influences the structure of the empirical model, which is explained in further detail in Chapter 5.

2.1 The Use of Attitudinal Data

McFadden (1986) provided early support for using attitudinal data in estimation of economic relationships. He considers a scenario in which the researcher has information on the product's characteristics, the individual's purchasing decision, and the individual's responses to attitudinal statements. These variables comprise a theoretical model in which unobserved personal characteristics interact with product attributes in the utility function to affect an individual's purchasing decision. The unobserved characteristics simultaneously influence the individual's responses to attitudinal questions. McFadden posits that jointly estimating purchasing decisions and attitudinal outcomes will produce more efficient parameter estimates. However, McFadden did not test this claim with an empirical application. My research builds on McFadden's theoretical model to confirm the efficiency gain hypothesis.

Cawley et al. (2004) use three waves of the National Longitudinal Survey of Youth to study the effects of body weight and cigarette prices on cigarette consumption. They estimate a discrete-time duration model to quantify the marginal effects of variables explaining smoking initiation. Within the model, the authors incorporate attitudinal responses to questions about the individual's weight status, thoughts on dieting, depression, and behavioral conduct as explanatory variables. They find significant relationships between these subjective responses and smoking initiation. However, as the central idea of my paper, I claim that these results might suffer from two types of bias: reverse causality and simultaneity. For example, there might exist reverse causality between smoking and dieting because one might contribute to the other or simultaneity between smoking and behavioral conduct because peer pressure, an unobserved factor, might simultaneously affect both outcomes. The authors acknowledge the endogeneity bias but lack strong instruments to perform a two stage regression.

Boxall and Adamowicz (2002) use attitudinal data to explain campsite decisions. The data include responses based on a 5-point Likert scale that ranges from "Strongly Disagree" to "Strongly Agree." For example, when given the statement "I go camping to relieve my tensions," the respondent may "Disagree," "Neither Disagree or Agree," etc. The authors regress a recreationist's campsite decision on the observable characteristics of each park, such as the user fee, the campsite type, and the campsite's level of development. They incorporate unobserved heterogeneity with a latent-segmentation model, in which the marginal effects of campsite characteristics are homogeneous among individuals in a particular segment but different across segments. An individual's responses to attitudinal statements (P_i) and demographic characteristics (X_i) influence her estimated latent-segment membership, and individuals belonging to latent segment μ_k have taste parameters α_k over campsite characteristics Z_c. That is,

$$\Pr\left(choice_i = c; \boldsymbol{\alpha}\right) = \sum_k \Pr(\mu = \mu_k | P_i, X_i) * \Pr\left(choice_i = c; \alpha_k | Z_c, \mu = \mu_k\right). \tag{2.1}$$

As a rebuttal to the previous paper, Morey et al. (2006) claim that the model (2.1) is misspecified, as it improperly incorporates unobserved heterogeneity by modeling latent-segment membership backwards. Morey et al. state that using attitudinal data to explain membership into a particular segment is not appropriate, as they instead "assume that [segment] membership is exogenous and that the probability of giving a particular level of response to an attitudinal question is a function of one's [segment]." To incorporate this intuition, the authors jointly model attitudinal responses with the demand equation and allow the unobserved heterogeneity to be correlated across equations. This specification expresses the idea that individuals from the same segment answer questions similarly, and not that individuals who happen to answer questions in the same manner must belong to the same segment.

As an empirical application for this 2006 publication, Morey et al. (2011) jointly estimate anglers' stated preferences among fishing grounds and their responses to attitudinal questions describing how bothered the anglers are by fish consumption advisories, how much they agree with particular "willingness-to-pay" statements, and how important certain fishing site attributes are to them. Their study offered anglers a choice set of hypothetical fishing grounds and recorded their stated preferences. The authors estimate a choice-only model and a joint choice-attitude model to show that the inclusion of attitudinal data does have an impact on the parameters of the choice probabilities, the parameters of the group membership probabilities, and the optimal number of latent groups to include in the final specification. Most importantly, the authors conclude that the joint model is small-sample more efficient than the choice-only model; the standard errors of the choice probability parameters are significantly smaller in the joint model. My research adds to their efforts by providing a theoretical motivation for including attitudinal responses, analyzing the econometric properties of the joint model, and using revealed preference data from a longitudinal survey.

Conti et al. (2010) employ attitudinal data to analyze the impact of childhood cognitive ability and psychosocial traits on later-in-life outcomes. The data contain measures related to the individual's academic ability, behavioral development, locus of control, and self-esteem at age ten and information on her education, health, and wage outcomes twenty years later.[1] The authors jointly estimate 126 psychometric measurements, one education decision, six health outcomes, and two wage equations, using nine explanatory variables as covariates (x_i) and ten latent factors (f_i) to represent unobserved heterogeneity. The specification for the model is:

$$\underset{135\times1}{y_i} = \underset{135\times9}{\beta} \times \underset{9\times1}{x_i} + \underset{135\times10}{\alpha} \times \underset{10\times1}{f_i} + \underset{135\times1}{\epsilon_i} . \tag{2.2}$$

Each latent factor is assumed to be normally distributed. Conditional on the covariates, the latent factors serve as the only source of unobserved correlation between the late-life outcomes

[1] Conti et al. (2010) use data from the 1970 British Cohort Study. All participants were born in 1970.

and these attitudinal data. The results report statistically significant parameter estimates for $\widehat{\alpha}$, suggesting that the correlation among unobserved factors that influence attitudinal and objective outcomes is substantial. While this specification is based on the same reasoning as the model presented in Chapter 5 of my paper, the authors do not focus on efficiency gains. In addition, my model relaxes the assumption of normality for the latent factors.

2.2 Assumptions About Unobserved Heterogeneity

Imposing a distributional assumption on the error terms is a convenient way to account for unobserved heterogeneity. However, Heckman and Singer (1984) demonstrate that an incorrect assumption could lead to biased parameters. Instead, the authors describe a semi-parametric specification that does not impose any distributional assumptions on the structure of unobserved heterogeneity. In an exploratory paper, Chintagunta, et al. (1991) compare several specifications of unobserved heterogeneity to the approach developed by Heckman and Singer (1984).[2] They use longitudinal data on household saltine cracker purchases to investigate consumers' heterogeneous brand preferences and conclude that the specified structure of heterogeneity plays an important role in determining the model's accuracy. Overall, the semi-parametric approach achieves the best model fit, as it generates more accurate predictions and outperforms alternative specifications on likelihood ratio tests.

Mroz and Guilkey (1995) and Mroz (1999) construct a Discrete-Factor Random Effects (DFRE) model that estimates the correlation across outcomes in a jointly-estimated set of equations. Like Heckman and Singer, the model approximates a discrete distribution for the unobserved heterogeneity by estimating the locations and probabilities for a finite number of mass points. As an extension, this DFRE model controls for both permanent and time-varying heterogeneity among the sample. Due to its flexibility in jointly modeling multiple outcomes, I use the DFRE estimation strategy for the Monte Carlo simulations (Chapter 6.2) and the empirical specification (Chapters 5 and 8) in this paper.

[2] The authors compare the semi-parametric model to specifications that incorporate heterogeneity through previous purchasing behavior, heterogeneous response parameters, gamma-distributed latent factors, and normally-distributed latent factors.

2.3 Models of Smoking Behavior

The dangers of developing smoking addictions at an early age have been widely publicized. Consequently, many health economists have conducted extensive research on the determinants of drug use among adolescents. During the 1990's, many state governments increased taxes on cigarettes in an attempt to discourage smoking initiation among adolescents. To measure the policy's effectiveness, DeCicca et al. (2002) estimate price elasticities resulting from the tax changes using the 1988-1992 National Education Longitudinal Study (NELS) dataset. The authors do not control for previous cigarette use as they limit the sample to individuals who are not smokers in the baseline year, but they do estimate a first-difference model to eliminate any permanent unobserved heterogeneity. With onset of smoking between eighth and twelfth grades as the dependent variable, their results suggest that the change in cigarette tax did not have a significant effect on smoking initiation. Instead, demographics, other state laws, and academic performance proved to be the significant influences on smoking behavior.

With the same NELS dataset, Gilleskie and Strumpf (2005) revisit the question using a model that includes previous period cigarette use. They add a time-invariant individual heterogeneity term via random effects to eliminate the endogeneity bias between consumption and lagged consumption resulting from permanent unobserved factors. The authors find small yet significant effects of previous smoking behavior in their unbiased coefficient estimates. Additionally, after incorporating these variables, estimates show that the tax increases were indeed effective in discouraging smoking initiation. They stress the importance of controlling for lagged behavior on top of permanent latent factors since "persistence cannot be fully explained by unobserved heterogeneity," as shown by the significant coefficients on lagged cigarette use intensities.

Arcidiacono et al. (2007) use data from the Health and Retirement Study to understand smoking and drinking behavior. Their structural approach relaxes the traditional assumption of a sample-wide discount factor (β). As a baseline, the authors construct a latent-segment model with a homogeneous β, which estimates an annual discount factor of 0.91. The next model allows the discount factor to vary based on the individual's latent segment. The unconstrained model yields estimates between 0.69 and 0.99 for different segments of the sample. This result suggests that an individual's discount factor is a significant source of unobserved heterogeneity in regards to decisions about cigarette and alcohol use. Their empirical findings provide motivation for the implications of the theoretical model of behavior in Chapter 3.

2.4 Contribution to the Literature

Overall, my paper expands upon the works of Morey et al. (2006) and Conti et al. (WP, 2010) by conducting a thorough investigation of the role of attitudinal data in estimation. First, my paper builds an **economic model of behavior** to motivate the inclusion of attitudinal data as an outcome rather than an exogenous variable within the empirical specification. Second, I provide an **econometric proof** to demonstrate how the jointly-estimated model improves the efficiency of the estimated parameters. Third, I construct a **Monte Carlo experiment** to validate this conclusion for small sample sizes. Finally, while the other papers only provide results from the jointly-estimated model, the **data application** in my paper compares the prediction accuracy of this specification against other commonly-used specifications.

Chapter 3

Theoretical Model of Consumer Behavior

Ultimately, this paper shows that the strength of the correlation among the unobserved factors that affect both the decision of interest (y) and attitudinal response (r) drives the efficiency gains for the model that jointly estimates y and r. The behavioral theory supplement to this paper, summarized here, demonstrates how three such unobserved factors representing the individual's *true* underlying preferences – the structural parameters $\boldsymbol{\mu}$ (a preference shifter), $\boldsymbol{\beta}$ (the discount factor), and $\boldsymbol{\alpha}$ (a subjective expectations parameter) – simultaneously impact the decision making process for y and the production of attitudinal response r.

In the theoretical model of consumer behavior, the inputs to the contemporaneous utility function for individual i in time period t are:

- y_{it} Outcome/decision of interest
- c_{it} Aggregate consumption good (excluding y_{it})
- X_{it} Exogenous factors
- Y_{it} Vector of variables describing an individual's decision history up to time t
- S_{it} Other information that impacts the individual's decision y_{it} and is affected by past decisions *(treated as a state variable)*
- ϵ_{it} Idiosyncratic utility shock

State variable S_{it} provides information about the individual's environment that impacts her decision y_{it}. In the model, the marginal utility from consuming y in period t varies based on the realization of S_{it}; thus, parameters of the contemporaneous utility function are indexed by S_{it}. More importantly, the current period decision, y_{it}, alters the probability distribution for the future period's state, S_{it+1}, which characterizes S_{it} as an endogenous variable. As an example from my data application, I would claim that smoking cigarettes ($y_{it} > 0$) becomes less desirable when the individual has trouble breathing (S_{it} = *adverse health shock*). Additionally, choosing to smoke more frequently can increase the probability of experiencing an adverse health shock in the next period.[1] In this model, the consumer does not know with certainty how her period t decision (y_{it}) affects her state realization next period (S_{it+1}). She forms a subjective expectation, represented by $\boldsymbol{\alpha}$, to approximate the impact of y_{it} on S_{it+1}. Without loss of generality, there are two possible states: $S_{it} \in \{0, 1\}$.

The individual's previous history of consumption represents another dynamic aspect of the decision making process. As an example, it has been shown that prolonged use of cigarettes is partly explained by habit formation. For this setting, variables for experience, duration, and cessation describe the history of previous decisions, Y_{it}.

3.1 Utility Function and Value Function

The contemporaneous utility function is given by:

$$U^{S_{it}}\left(y_{it}, c_{it}; \boldsymbol{\mu}, X_{it}, Y_{it}, \epsilon_{it}^{y}\right). \tag{3.1}$$

The individual spends her income on y_{it} and general consumption goods c_{it}. The primitive parameter $\boldsymbol{\mu}$ represents a preference shifter for consumption of y_{it}, and I assume that the marginal utility of consuming good y is a function of $\boldsymbol{\mu}$, among several other factors. In traditional economic models, $\boldsymbol{\mu}$ might enter the utility function as a risk aversion parameter or as a taste parameter. In this dynamic optimization problem, the consumer's current period decisions affect future realizations of history Y_{it} and state S_{it} variables. The individual is forward looking in that she takes into account the discounted expected utility from future time periods when making a decision for the current period. The discount factor $\boldsymbol{\beta}$, which describes how the individual values future streams of utility, is another behavioral primitive of the optimization problem. For each period, the individual considers the contemporaneous

[1]Other decision-state pairs exhibiting this relationship include weight loss food consumption and weight, cleaning supply usage and household cleanliness, and dietary supplement/steroid usage and strength.

utility from that period's decision and the effects of this decision on the discounted present value of future utility flows.

I assume that, in period t, the individual does not know the marginal effect of her decision y_{it} on next period's state realization S_{it+1} and must generate a subjective expectation over this parameter. Her subjective-expectations operator, $\boldsymbol{\alpha}$, represents the individual's belief of how her decision impacts future state transitions and enters the expected state transition probabilities $\pi_{0it}\left(\boldsymbol{\alpha}, y_{it}, X_{it}, Y_{it}\right) = \mathbb{E}\left[\Pr\left(S_{it+1}=0\right)\right]$ and $\pi_{1it}\left(\boldsymbol{\alpha}, y_{it}, X_{it}, Y_{it}\right) = \mathbb{E}\left[\Pr\left(S_{it+1}=1\right)\right]$. Her conditional (on next period's state realization S_{it+1}) maximum expected future lifetime utilities are represented by $V^0\left(\bullet\right)$ and $V^1\left(\bullet\right)$. Conditional on the current period state realization $S_{it}=S$, the value function for alternative $y_{it}=y$ is given by:

$$
\begin{aligned}
V_y^S\left(\boldsymbol{\mu}, \boldsymbol{\beta}, \boldsymbol{\alpha}, X_{it}, Y_{it}, \epsilon_{it}^{yS}\right) = U^{S_{it}=S}&(y_{it}=y, \left[I_{it}-p_t y\right]; \boldsymbol{\mu}, X_{it}, Y_{it}, \epsilon_{it}^{yS}) \\
&+\boldsymbol{\beta} * \pi_1(\boldsymbol{\alpha}, y_{it}, X_{it}, Y_{it}) * V^1\left(\boldsymbol{\mu}, \boldsymbol{\beta}, \boldsymbol{\alpha}, X_{it+1}, Y_{it+1}\right) \\
&+\boldsymbol{\beta} * \pi_0(\boldsymbol{\alpha}, y_{it}, X_{it}, Y_{it}) * V^0\left(\boldsymbol{\mu}, \boldsymbol{\beta}, \boldsymbol{\alpha}, X_{it+1}, Y_{it+1}\right).
\end{aligned} \tag{3.2}
$$

Here, general consumption is inferred from income I_{it} and the price of y_{it}, p_t, as $c_{it} = I_{it} - p_t y_{it}$.

In summary, the relevant primitives of the optimization problem that will drive the unobserved correlation across equations enter the model through the:

- Preference shifters, $\boldsymbol{\mu}$
- Time preference parameter (discount factor), $\boldsymbol{\beta}$
- Subjective expectations over future states, $\boldsymbol{\alpha}$.

3.2 Optimal Decision Rules for Decision y_{it}

Using the value function specification in (3.2), the individual chooses at each period t the alternative that maximizes her remaining lifetime utility. The discrete-choice framework uses choice probabilities to describe the probability that outcome $y_{it}=y$ occurs. That is,

$$
\begin{aligned}
&\Pr\left(y_{it}=y|\boldsymbol{\mu},\boldsymbol{\beta},\boldsymbol{\alpha},X_{it},Y_{it},S_{it}=S,\epsilon_{it}\right)\\
&\qquad=\Pr\left(V_y^S\left(\boldsymbol{\mu},\boldsymbol{\beta},\boldsymbol{\alpha},X_{it},Y_{it},\epsilon_{it}^{yS}\right)\geq V_{\tilde{y}}^S\left(\boldsymbol{\mu},\boldsymbol{\beta},\boldsymbol{\alpha},X_{it},Y_{it},\epsilon_{it}^{\tilde{y}S}\right),\forall\ \tilde{y}\neq y\right).
\end{aligned}
\tag{3.3}
$$

Clearly, the choice probabilities are functions of primitive parameters $\boldsymbol{\mu}$, $\boldsymbol{\beta}$, and $\boldsymbol{\alpha}$ because they define the optimization problem. A more rigorous proof in the **Appendix** derives the optimality conditions to show this relationship more explicitly.

3.3 Production of Attitudinal Responses r_{it}

Contrary to *decisions* y_{it}, attitudinal responses are *outcomes* that are influenced by several factors of the individual's environment. She, herself, does not choose her attitudes; she merely reports the outcome that occurred. Similar to the value function of equation (3.3), the attitudinal index function $A(\bullet)$ describes how the individual's response to an attitudinal question is produced. The probability of giving a particular response is a function of personal characteristics, primitive parameters, and other unobserved factors. That is,

$$
\begin{aligned}
&\Pr\left(r_{it}=r|\boldsymbol{\mu},\boldsymbol{\beta},\boldsymbol{\alpha},X_{it},Y_{it},S_{it}=S,\eta_{it}\right)\\
&\qquad=\Pr\left(A_r^S\left(\boldsymbol{\mu},\boldsymbol{\beta},\boldsymbol{\alpha},X_{it},Y_{it},\eta_{it}^{rS}\right)\geq A_{\tilde{r}}^S\left(\boldsymbol{\mu},\boldsymbol{\beta},\boldsymbol{\alpha},X_{it},Y_{it},\eta_{it}^{\tilde{r}S}\right),\forall\ \tilde{r}\neq r\right).
\end{aligned}
\tag{3.4}
$$

The factors that influence responses to attitudinal questions can be grouped into four categories:

1. **Exogenous characteristics $\boldsymbol{X}_{it}$, history vector $\boldsymbol{Y}_{it}$, and state $\boldsymbol{S}_{it}$** explain systematic differences in the way individuals generate attitudinal responses.

2. In deciding which attitudinal response questions are relevant to the demand behavior of interest, I select only those questions that can conceivably be affected by **primitive parameters** such as preference shifter $\boldsymbol{\mu}$, the discount factor $\boldsymbol{\beta}$, and the subjective-expectations operator $\boldsymbol{\alpha}$.

3. No survey can capture everything that affects a person's behavior; there are several **unobserved factors and events** that influence the individual's attitudinal responses on the survey.

4. Attitudes and feelings are intangible and difficult for the respondent to quantify on a survey form. Reporting opinions based on a 5-point Likert scale presents the possibility of **measurement error** in reporting her true beliefs.

Factors described in categories 3 and 4 comprise the composite error term η_{it}. Factors described in categories 1, 2, and 3 might simultaneously influence the outcome y_{it}. The important latent correlation between y_{it} and r_{it} that justifies the joint estimation proposed by this paper is driven by factors found in categories 2 and 3, as ultimately these data will be unobserved to the econometrician in the empirical model. The implications of this effect are explained in greater detail in the next section.

Chapter 4

Implications for the Empirical Model

For the data application, I do not include an endogenous state variable in order to keep the model more tractable.[1] Thus, the model comprises equations (3.3) and (3.4):

$$\begin{aligned}
&\Pr\left(y_{it}=y|\boldsymbol{\mu},\boldsymbol{\beta},\boldsymbol{\alpha},X_{it},Y_{it},\epsilon_{it}\right)\\
&\qquad=\Pr\left(V_y\left(\boldsymbol{\mu},\boldsymbol{\beta},\boldsymbol{\alpha},X_{it},Y_{it},\epsilon_{it}^{y}\right)\geq V_{\tilde{y}}\left(\boldsymbol{\mu},\boldsymbol{\beta},\boldsymbol{\alpha},X_{it},Y_{it},\epsilon_{it}^{\tilde{y}}\right),\forall\ \tilde{y}\neq y\right)\\
&\Pr\left(r_{it}=r|\boldsymbol{\mu},\boldsymbol{\beta},\boldsymbol{\alpha},X_{it},Y_{it},\eta_{it}\right)\\
&\qquad=\Pr\left(A_r\left(\boldsymbol{\mu},\boldsymbol{\beta},\boldsymbol{\alpha},X_{it},Y_{it},\eta_{it}^{r}\right)\geq A_{\tilde{r}}\left(\boldsymbol{\mu},\boldsymbol{\beta},\boldsymbol{\alpha},X_{it},Y_{it},\eta_{it}^{\tilde{r}}\right),\forall\ \tilde{r}\neq r\right)
\end{aligned}\tag{4.1}$$

I use Taylor Series approximations to the alternative-specific value and attitudinal index functions to obtain reduced form expressions for the demand function for y_{it} and production function for r_{it}. The following outcome probabilities represent the linear approximations of functions V_y and A_r, in which $\{\gamma,\phi\}$ and $\{\omega_{it},\tau_{it}\}$ are the coefficients and error terms of the Taylor Series approximations:

$$\begin{aligned}
&\Pr\left(y_{it}=y|\boldsymbol{\mu},\boldsymbol{\beta},\boldsymbol{\alpha},X_{it},Y_{it},\epsilon_{it}\right)\\
&\qquad=\Pr\left(V_y\left(\boldsymbol{\mu},\boldsymbol{\beta},\boldsymbol{\alpha},X_{it},Y_{it},\epsilon_{it}^{y}\right)\geq V_{\tilde{y}}\left(\boldsymbol{\mu},\boldsymbol{\beta},\boldsymbol{\alpha},X_{it},Y_{it},\epsilon_{it}^{\tilde{y}}\right),\forall\ \tilde{y}\neq y\right)\\
&\qquad\approx\Pr\left(X_{it}\gamma_X^{y}+Y_{it}\gamma_Y^{y}+\omega_{it}^{y}\geq X_{it}\gamma_X^{\tilde{y}}+Y_{it}\gamma_Y^{\tilde{y}}+\omega_{it}^{\tilde{y}},\forall\ \tilde{y}\neq y\right)\\
&\qquad=\Pr\left(\omega_{it}^{y}-\omega_{it}^{\tilde{y}}\geq X_{it}\left(\gamma_X^{\tilde{y}}-\gamma_X^{y}\right)+Y_{it}\left(\gamma_Y^{\tilde{y}}-\gamma_Y^{y}\right),\forall\ \tilde{y}\neq y\right)\\
&\qquad\equiv\Pr\left(y_{it}=y|X_{it},Y_{it}\right)
\end{aligned}\tag{4.2}$$

[1] In the **Appendix**, I outline a specification that estimates the transition equation for an endogenous state variable.

$$
\begin{aligned}
&\Pr\left(r_{it}=r|\boldsymbol{\mu},\boldsymbol{\beta},\boldsymbol{\alpha},X_{it},Y_{it},\eta_{it}\right)\\
&\qquad=\Pr\left(A_r\left(\boldsymbol{\mu},\boldsymbol{\beta},\boldsymbol{\alpha},X_{it},Y_{it},\eta_{it}^r\right)\geq A_{\tilde{r}}\left(\boldsymbol{\mu},\boldsymbol{\beta},\boldsymbol{\alpha},X_{it},Y_{it},\eta_{it}^{\tilde{r}}\right),\forall\ \tilde{r}\neq r\right)\\
&\qquad\approx\Pr\left(X_{it}\phi_X^r+Y_{it}\phi_Y^r+\tau_{it}^r\geq X_{it}\phi_X^{\tilde{r}}+Y_{it}\phi_Y^{\tilde{r}}+\tau_{it}^{\tilde{r}},\forall\ \tilde{r}\neq r\right)\\
&\qquad=\Pr\left(\tau_{it}^r-\tau_{it}^{\tilde{r}}\geq X_{it}\left(\phi_X^{\tilde{r}}-\phi_X^r\right)+Y_{it}\left(\phi_Y^{\tilde{r}}-\phi_Y^r\right),\forall\ \tilde{r}\neq r\right)\\
&\qquad\equiv\Pr\left(r_{it}=r|X_{it},Y_{it}\right)\qquad.
\end{aligned}
\tag{4.3}
$$

The approximation approach is common among empirical economic research because it reduces the computation time while retaining the ability to predict outcomes and produce policy simulations. This method, however, does not recover the structural parameters that shape an individual's decision making problem. For instance, in equations (4.2) and (4.3), the total effects of the explanatory variables on the outcome probabilities are aggregated into the response parameter vectors $\{\gamma_X,\gamma_Y\}$ and $\{\phi_X,\phi_Y\}$, which are functions of the underlying structural parameters $\boldsymbol{\mu}$, $\boldsymbol{\beta}$, and $\boldsymbol{\alpha}$.

The theoretical model in Chapter 3 suggests that error terms ω_{it} and τ_{it} are correlated, and the econometric analysis discussed later in Chapter 6 shows that jointly estimating outcomes y_{it} and r_{it} produces more efficient parameter estimates. A random effects model accommodates these considerations by incorporating latent factors λ and ν_t into each equation to account for the shared unobserved heterogeneity. The error terms are decomposed into three components:

$$
\begin{aligned}
\omega_{it}^y&=\lambda^y+\nu_t^y+\xi_{it}^y\\
\tau_{it}^r&=\lambda^r+\nu_t^r+\zeta_{it}^r
\end{aligned}
\tag{4.4}
$$

Here, $\mathbb{E}\left[\lambda^y,\lambda^r\right]\neq\{0,0\}$, $\mathbb{E}\left[\nu_t^y,\nu_t^r\right]\neq\{0,0\}$, and ξ_{it} and ζ_{it} are *iid* serially uncorrelated errors. An assumption that the idiosyncratic shocks ξ_{it} and ζ_{it} follow independent GEV distributions creates a closed form expression for the choice probabilities, given by:

$$
\Pr\left(y_{it}=y|X_{it},Y_{it},\lambda^y,\nu_t^y\right)=\frac{\exp\left(X_{it}\gamma_X^y+Y_{it}\gamma_Y^y+\lambda^y+\nu_t^y\right)}{\sum_{\tilde{y}}\exp\left(X_{it}\gamma_X^{\tilde{y}}+Y_{it}\gamma_Y^{\tilde{y}}+\lambda^{\tilde{y}}+\nu_t^{\tilde{y}}\right)}
\tag{4.5}
$$

$$
\Pr\left(r_{it}=r|X_{it},Y_{it},\lambda^r,\nu_t^r\right)=\frac{\exp\left(X_{it}\phi_X^r+Y_{it}\phi_Y^r+\lambda^r+\nu_t^r\right)}{\sum_{\tilde{r}}\exp\left(X_{it}\phi_X^{\tilde{r}}+Y_{it}\phi_Y^{\tilde{r}}+\lambda^{\tilde{r}}+\nu_t^{\tilde{r}}\right)}
\tag{4.6}
$$

Further details on estimation are explained in Chapter 5.

4.1 Comparison: Attitudinal Responses are Treated as Exogenous Variables

In the recent economic literature, many empirical specifications do include data on the individual's attitudes. However, attitudinal responses are often treated as explanatory variables in the choice probabilities for y_{it}, where

$$\begin{aligned} &\Pr\left(y_{it} = y | X_{it}, Y_{it}, r_{it}\right) \\ &\quad = \Pr\left(\omega_{it}^{y} - \omega_{it}^{\tilde{y}} \geq X_{it}\left(\gamma_X^{\tilde{y}} - \gamma_X^{y}\right) + Y_{it}\left(\gamma_Y^{\tilde{y}} - \gamma_Y^{y}\right) + r_{it}\left(\gamma_r^{\tilde{y}} - \gamma_r^{y}\right), \forall\ \tilde{y} \neq y\right). \end{aligned} \tag{4.7}$$

Here, r_{it} enters the problem as an exogenous right-hand-side variable. The conclusions from the theoretical model of behavior in Chapter 3.3 contradict this assumption because unobserved factors ω_{it} simultaneously affect attitudinal response r_{it}. A biased estimate for response parameter γ_r and inefficient estimates for all other parameters result from this specification. Furthermore, significant multicollinearity between the endogenous response variable r_{it} and other explanatory variables could bias response parameters γ_X and γ_Y as well. Incorporating latent factors does not eliminate the endogeneity bias since the initial specification (4.7) assumes independence between the explanatory variables and *all* components of error term ω_{it}.

Attitudinal data have also been used to control for unobserved heterogeneity in latent-segment models (Boxall and Adamowicz, 2002). The model estimates response parameters for different segments of the sample, and each individual in the sample is placed into a segment based on her attitudinal responses. However, this model is misspecified as it imposes an incorrect direction of causality. Instead, my model assumes that all individuals from a particular segment respond to attitudinal questions in the same manner, with an idiosyncratic shock to explain the differences, which emphasizes that the individual's type is the exogenous personality trait that influences her responses on the survey.

Chapter 5

Empirical Specification

In the data application, the preferred model jointly estimates the choice probability for y (4.5) with the response probabilities for Q attitudinal questions of the form (4.6). The possible response categories for y_{it} are $\{0, \ldots, \mathrm{Y}\}$ and for attitudinal question r_q are $\{0, \ldots, \mathrm{R}_q\}$. For identification, the model normalizes parameters with respect to alternative 0.[1] The likelihood contribution of the period t outcomes for individual i, conditional on latent factors λ and ν_t, equals:

$$\begin{aligned} &\ell\left(y_{it}, \mathcal{R}_{it} | \boldsymbol{z}_{it}, \lambda, \nu_t; \Theta\right) \\ &\quad = \prod_{y=0}^{\mathrm{Y}} \Pr\left(y_{it} = y | \boldsymbol{z}_{it}, \lambda, \nu_t\right)^{\mathbb{1}(y_{it}=y)} * \prod_{q=1}^{Q} \prod_{r_q=0}^{\mathrm{R}_q} \Pr\left(r_{qit} = r_q | \boldsymbol{z}_{it}, \lambda, \nu_t\right)^{\mathbb{1}(r_{qit}=r)}. \end{aligned} \tag{5.1}$$

For brevity, I represent attitudinal responses $\{r_{1it}, \ldots, r_{Qit}\}$ with vector $\mathcal{R}_{it}$, condense characteristics $\{X_{it}, Y_{it}\}$ into $\boldsymbol{z}_{it}$, consolidate response parameters $\{\gamma, \phi\}$ into Θ, and index attitudinal questions by q. The Discrete-Factor Random Effects (DFRE) model approximates the latent-factor distributions by estimating K mass points for the vector λ and L mass points for the vector ν_t, and the dimensional size of each mass point equals the number of equation-alternative pairs in the model. This method associates probabilities $\{\mathfrak{p}_k\}_{k=1}^{K}$ to mass points $\{\lambda_k\}_{k=1}^{K}$ and probabilities $\{\mathfrak{q}_l\}_{l=1}^{L}$ to mass points $\{\nu_{tl}\}_{l=1}^{L}$. For illustration, the latent factor parameterization for mass point λ_1 is summarized by

$$\lambda_1 = \{\underbrace{\lambda_1^{y1}, \ldots, \lambda_1^{y\mathrm{Y}}}_{\substack{\text{equation for } y, \\ \text{alt's 1,...,Y}}}, \underbrace{\lambda_1^{11}, \ldots, \lambda_1^{1\mathrm{R}_1}}_{\substack{\text{equation for } r_1, \\ \text{alt's 1,...,R}_1}}, \ldots, \underbrace{\lambda_1^{Q1}, \ldots, \lambda_1^{Q\mathrm{R}_Q}}_{\substack{\text{equation for } r_Q, \\ \text{alt's 1,...,R}_Q}}\} \qquad \{\mathfrak{p}_1\} = \Pr\left(\lambda = \lambda_1\right). \tag{5.2}$$

---

[1] For alternative 0, $\Pr\left(y_{it} = 0 | X_{it}, Y_{it}, \lambda, \nu_t\right) = 1 - \Sigma_{y=1}^{\mathrm{Y}} \Pr\left(y_{it} = y | X_{it}, Y_{it}, \lambda, \nu_t\right)$ and likewise for $\Pr\left(r_{qit} = 0 | \bullet\right)$.

The unconditional likelihood contribution for individual i becomes:

$$\mathcal{L}\left(\{y_{it},\mathcal{R}_{it}\}_{t=1}^{T} \,|\, \{\boldsymbol{z}_{it}\}_{t=1}^{T}; \Theta, \mathfrak{p}, \mathfrak{q}\right) = \sum_{k=1}^{K} \mathfrak{p}_k \prod_{t=1}^{T} \sum_{l=1}^{L} \mathfrak{q}_l * \ell\left(y_{it},\mathcal{R}_{it}|\boldsymbol{z}_{it},\lambda_k,\nu_{tl};\Theta\right). \tag{5.3}$$

Aggregating across all individuals in the sample creates the empirical likelihood function,

$$\mathfrak{L}\left(\boldsymbol{y},\boldsymbol{\mathcal{R}}|\boldsymbol{Z};\Theta,\mathfrak{p},\mathfrak{q}\right) = \prod_{i=1}^{N} \mathcal{L}\left(\{y_{it},\mathcal{R}_{it}\}_{t=1}^{T} \,|\, \{\boldsymbol{z}_{it}\}_{t=1}^{T}; \Theta, \mathfrak{p}, \mathfrak{q}\right). \tag{5.4}$$

Chapter 6

Econometric Motivation

It seems reasonable that pairing an individual's personality and opinions with her purchasing decision will provide a better understanding of what motivates her to use that product. The challenge, then, is fitting these attitudinal responses into an empirical model in a way that both supports the theory of decision making and improves estimation. In Chapter 3.3, I develop a jointly-estimated model of outcomes and attitudinal responses that includes an identical set of explanatory variables in each equation. A seemingly unrelated regression (SUR) model, introduced by Zellner (1962, 1963), accommodates this specification. By accounting for the correlation among unobserved factors affecting attitudinal and purchasing outcomes, the SUR model produces parameter estimates with smaller standard errors. However, there is one caveat: a SUR model of *continuous* dependent variables that uses the same set of covariates in each equation generates identical results to an OLS model, without improving efficiency. This is the case for the empirical model in this paper, except that the outcomes are *categorical*. The econometric theory supplement to this paper adapts an econometric proof from McCullagh and Nelder (1989) to show that efficiency can still be gained when jointly estimating two or more correlated categorical dependent variables. Its conclusions are summarized in Section 6.1. Section 6.2 presents Monte Carlo simulations to support the theory.

6.1 Fisher Information Matrix

I examine a case of two binary random variables, $y \in \{1, 2\}$ and $r \in \{1, 2\}$, in which y is the outcome of interest and r is the response to an attitudinal question.[1] Diagonal elements of the inverse of the Fisher information matrix are used to find the asymptotic variance of the parameters in the choice probabilities. Hence, I present the Fisher information matrices for two logit models in order to show that the standard errors of the estimated parameters are smaller when the unobserved factors are allowed to be Correlated across equations in **Model C** than when they are assumed to be Independent in **Model I**. Ultimately, the degree of correlation between the unobserved factors that affect y and the unobserved factors that affect r determines the efficiency gains in **Model C** over **Model I**.

- **Model C** jointly models the outcome, attitudinal response, and correlation, in which the joint probability, $\Pr\left(y = j, r = k | X; \beta^C\right)$, cannot be separated into marginal probabilities without modeling the correlation coefficient.

- **Model I** assumes conditional independence between the outcome and attitudinal response, such that:

$$\Pr\left(y = j, r = k | X; \beta^I\right) = \Pr\left(y = j | X; \beta^I_y\right) * \Pr\left(r = k | X; \beta^I_r\right).$$

Due to the independence assumption in **Model I**, simply estimating $\Pr\left(y | X; \beta^I_y\right)$ by itself will produce an identical point estimate and variance for response parameter β^I_y. Thus, **Model I** is equivalent to a standard logit specification that estimates only the choice probability for y, which is the case when attitudinal information is either ignored or not collected. For **Model C**, a bivariate logistic model accurately captures the relationship between two correlated binary random variables (y, r).[2] For outcome y, outcome r, and a set of explanatory variables X, the parameterizations for the log-odds ratios of the jointly-estimated model are:

[1]McCullagh and Nelder (1989) explain how this case can easily be extended to accommodate categorical variables with more than two alternatives.

[2]Ultimately, I use a logistic model in my empirical model. A similar proof, used in Meng and Schmidt (1985), derives the Fisher information matrix for a bivariate probit model, which shows the same results as those presented here.

$$\log\left(\frac{\Pr(y=1|X)}{\Pr(y=2|X)}\right) = \beta_y^C X,$$
$$\log\left(\frac{\Pr(r=1|X)}{\Pr(r=2|X)}\right) = \beta_r^C X, \tag{6.1}$$
$$\text{and } \log\left(\frac{\Pr(y=1,r=1|X)*\Pr(y=2,r=2|X)}{\Pr(y=1,r=2|X)*\Pr(y=2,r=1|X)}\right) = \beta_{yr}^C.$$

The first two equations estimate response parameters β_y^C and β_r^C while the third equation estimates the correlation coefficient, β_{yr}^C, between the error terms.

My primary focus turns to the asymptotic properties of β_y, the response parameter for X on the outcome of interest, for each model. The proof in McCullagh and Nelder (1989) shows that both estimators β_y^C and β_y^I are asymptotically unbiased, but the estimator β_y^C is more efficient. To compare the variances of β_y^C and β_y^I, the econometrician only needs to analyze the second principal minor of the Fisher information matrix for **Model C** and the first principal minor of the matrix for **Model I**:[3]

$$\mathcal{F}_{2\text{nd}}^{\boldsymbol{M(C)}} = \begin{array}{cc} & \begin{array}{cc} \beta_y^C & \beta_r^C \end{array} \\ \begin{array}{c} \beta_y^C \\ \beta_r^C \end{array} & \begin{pmatrix} \boldsymbol{X}'diag\left\{\frac{V_y}{\Delta}\right\}\boldsymbol{X} & \boldsymbol{X}'diag\left\{\frac{\Delta_\pi}{\Delta}\right\}\boldsymbol{X} \\ \boldsymbol{X}'diag\left\{\frac{\Delta_\pi}{\Delta}\right\}\boldsymbol{X} & \boldsymbol{X}'diag\left\{\frac{V_r}{\Delta}\right\}\boldsymbol{X} \end{pmatrix} \end{array}. \tag{6.2}$$

$$\mathcal{F}_{1\text{st}}^{\boldsymbol{M(I)}} = \begin{array}{cc} & \beta_y^I \\ \beta_y^I & \left(\boldsymbol{X}'diag\left\{V_y\right\}\boldsymbol{X}\right) \end{array} \tag{6.3}$$

The terms V_y, V_r, V_{yr}, Δ, and Δ_π represent functions of joint probabilities $\Pr(y=j, r=k|X)$, and the operator $diag\{\bullet\}$ transforms the $N \times 1$ vector in brackets into a diagonal matrix of rank N.

Equations (6.2) and (6.3) show that the value of Δ plays an important role in determining the efficiency gains of **Model C** over **Model I**. When the error terms are independent, then $\Delta = 1$ and $\Delta_\pi = 0$. Under this condition, the variances of β_y^C and β_y^I will be identical, and the jointly-estimated system of outcomes provides no benefits. However, if the error terms are

[3]Some rows and columns of each matrix can be ignored as they do not affect the variance of β_y, found in the inverse Fisher information matrices.

correlated, then $\Delta < 1$, and the $(1,1)$ element of $\mathcal{F}^{\boldsymbol{M(C)}}$ is *greater* than the $(1,1)$ element of $\mathcal{F}^{\boldsymbol{M(I)}}$. When the matrices are inverted to attain the variances, this relationship is reversed, which means the variance of β_y^C will be *smaller* than the variance of β_y^I.

6.2 Monte Carlo Simulations

This Monte Carlo experiment uses a simple data generating process to compare five empirical models, ultimately showing that the jointly-estimated model improves the small-sample accuracy and efficiency of the estimated parameters when significant correlation exists among the unobserved factors that influence the outcome y and the attitudinal response r. To measure these gains under different environments, 96 unique data specifications are generated by specifying the number of individuals, number of time periods, latent-factor distribution, and strength of correlation between error terms across outcomes.[4] The eight individual-period pairs include either 500 or 1,000 individuals for 1, 3, 5, or 10 time periods.

Table 6.1: Monte Carlo Specification Options: Individual-Period Pairs

(N, T)	# Periods = 1	# Periods = 3	# Periods = 5	# Periods = 10
# Individuals = 500	$(500, 1)$	$(500, 3)$	$(500, 10)$	$(500, 10)$
# Individuals = 1000	$(1000, 1)$	$(1000, 3)$	$(1000, 10)$	$(1000, 10)$

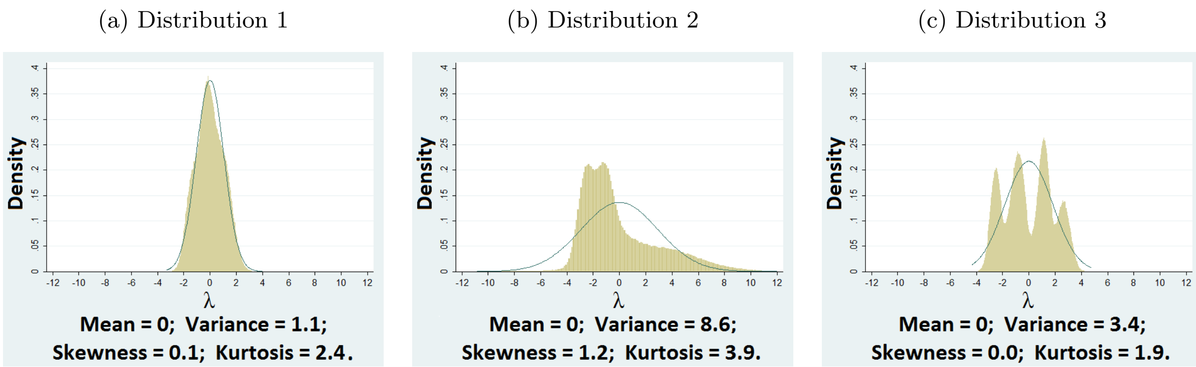
6.2.1 Unobserved Heterogeneity

The econometric theory supplement shows that the efficiency gain increases as the correlation among error terms grows stronger. To test this theory, the Monte Carlo data generating processes allow for differences in the error correlation structure. One time-invariant latent factor enters the data generating processes for outcomes y and r. To relax the assumption that the latent terms must follow a conventional distribution, each individual's latent factor is independently drawn from a mixture of normal distributions. For robustness, three such mixture distributions, shown in Figure 6.1, are considered. Distribution 1 is very close to the standard normal density. Distribution 2 is skewed to the left and has a heavy right tail. Distribution 3

[4]There are two options for the number of individuals, four options for the number of time periods, three latent-factor distributions, and four options for the factor loading on the latent factor.

contains four distinct high-density areas, but its variance is smaller than that of Distribution 2. A normal-density bell curve overlays each graph in Figure 6.1 for comparison.

Figure 6.1: Distribution of latent factor λ



In each outcome's data generating process (DGP), the composite error term is composed of an additional independently- and identically-distributed (*iid*) generalized extreme value (*GEV*) disturbance (*explained in the following section*) and the composite latent-factor term. The latent term λ_i influences both outcomes and enters both DGP's additively. The magnitude of latent factor λ_i in the DGP for outcome y is constant across specifications. This magnitude in the DGP for outcome r, represented by ρ^r, varies from zero to 0.9 ($\rho^r = \{0, 0.3, 0.6, 0.9\}$). There is no correlation across error terms for specifications in which $\rho^r = 0$, but this correlation strengthens as ρ^r increases. This relationship is illustrated in the DGP summary of the following section.

6.2.2 Data Generation for Longitudinal Specifications

An initial condition, y_{i0}, is estimated for longitudinal specifications in which $T > 1$. Conditional on an exogenous characteristic (W_{i0}) and the latent factor (λ_i), dichotomous outcome y_{i0} follows an *iid GEV* distribution, which creates a closed-form expression for the outcome probabilities. For all outcomes, I construct a latent index, compute the logistic probability, and draw a random variable Ψ from a standard uniform distribution, $U[0, 1]$. The model produces the outcome $y_{i0} = 1$ if the logit probability exceeds random variable Ψ and the outcome $y_{i0} = 0$ otherwise. The data generation process for y_{i0} is summarized by the following:

1. Draw $W_{i0} \sim \mathcal{N}(0, 36)$ and λ_i from the specified distribution.
2. Construct latent index $y^*_{i0} = \beta^{y_0}_0 + \beta^{y_0}_1 W_{i0} + \lambda_i$.
3. Draw $\Psi^y_{i0} \sim U[0, 1]$.
4. Generate outcome $y_{i0} = \mathbb{1}\left(\frac{\exp(y^*_{i0})}{1+\exp(y^*_{i0})} \geq \Psi^y_{i0}\right)$.

The outcomes in the remaining $T-1$ periods are generated through a similar process. To incorporate persistence in purchasing behavior, the previous period outcome y_{it-1} influences current period outcome y_{it}. As a result, an empirical specification that does not account for the serial autocorrelation caused by time-invariant latent factor λ_i will produce a biased estimate for the marginal effect of y_{it-1} on y_{it}. Previous period outcomes y_{it-1} and r_{it-1} are not included in the data generation process of current period attitudinal responses r_{it} so that the model conforms to the assumptions of the theoretical model and empirical specification in this paper.[5] A time-varying exogenous characteristic, X_{it}, influences both outcomes y_{it} and r_{it}. The entire data generating process is summarized below:

$$W_{i0} \sim \mathcal{N}(0, 36) \text{ and } X_{it} \sim \mathcal{N}(0, 36) \text{ for } t = 1, \ldots, T-1.$$

λ_i is drawn from the specified latent-factor distribution.

$\Psi^y_{i0}, \Psi^y_{i1}, \Psi^r_{i1}, \ldots, \Psi^y_{iT-1}$, and Ψ^r_{iT-1} are drawn independently from $U[0, 1]$.

$$y_{i0} = \mathbb{1}\left(\frac{\exp(y^*_{i0})}{1+\exp(y^*_{i0})} \geq \Psi^y_{i0}\right), \text{ where } y^*_{i0} = \beta^{y_0}_0 + \beta^{y_0}_1 W_{i0} + \lambda_i.$$

$$y_{it} = \mathbb{1}\left(\frac{\exp(y^*_{it})}{1+\exp(y^*_{it})} \geq \Psi^y_{it}\right), \text{ where } y^*_{it} = \beta^y_0 + \beta^y_1 y_{it-1} + \beta^y_2 X_{it} + \lambda_i, \quad \text{for } t = 1, \ldots, T-1$$

$$r_{it} = \mathbb{1}\left(\frac{\exp(r^*_{it})}{1+\exp(r^*_{it})} \geq \Psi^r_{it}\right), \text{ where } r^*_{it} = \beta^r_0 \qquad + \beta^r_2 X_{it} + \rho^r \lambda_i, \text{ for } t = 1, \ldots, T-1.$$

The coefficients of the model and distributions of exogenous characteristics are chosen so that latent factor λ_i does not overpower the influence of observed explanatory variables on outcomes y and r (*explained in Section 6.2.4*). For this Monte Carlo experiment, $\beta^{y_0}_0 = 1$, $\beta^{y_0}_1 = -0.5$, $\beta^y_0 = -1$, $\beta^y_1 = 0.75$, $\beta^y_2 = -0.5$, $\beta^r_0 = -1$, and $\beta^r_2 = 0.5$. For illustration, the following steps

[5] Results are available for a specification in which y_{it-1} is also allowed to influence r_{it}. These results show no difference in model performance.

outline the data generation process for a specification in which $N = 500$, $T = 5$, latent-factor distribution 2 is used, and $\rho^r = 0.9$:

1. For 500 individuals, draw $W_{i0} \sim \mathcal{N}(0, 36)$ and λ_i from latent-factor distribution 2.
2. Construct latent index $y^*_{i0} = 1 - 0.5\, W_{i0} + \lambda_i$.
3. Draw $\Psi^y_{i0} \sim U[0, 1]$.
4. Generate outcome $y_{i0} = \mathbb{1}\left(\frac{\exp\left(y^*_{i0}\right)}{1+\exp\left(y^*_{i0}\right)} \geq \Psi^y_{i0}\right)$.
5. Draw $X_{i1} \sim \mathcal{N}(0, 36)$.
6. Construct latent indeces $y^*_{i1} = -1 + 0.75\, y_{i0} - 0.5\, X_{i1} + \lambda_i$ and $r^*_{i1} = -1 + 0.5\, X_{i1} + 0.9\, \lambda_i$.
7. Draw $\Psi^y_{i1} \sim U[0, 1]$ and $\Psi^r_{i1} \sim U[0, 1]$.
8. Generate outcomes $y_{i1} = \mathbb{1}\left(\frac{\exp\left(y^*_{i1}\right)}{1+\exp\left(y^*_{i1}\right)} \geq \Psi^y_{i1}\right)$ and $r_{i1} = \mathbb{1}\left(\frac{\exp\left(r^*_{i1}\right)}{1+\exp\left(r^*_{i1}\right)} \geq \Psi^r_{i1}\right)$.
9. Repeat steps 5. through 8. for time periods 2, 3, and 4.

6.2.3 Data Generation for Cross-Sectional Specifications

For the cross-sectional specifications, only one outcome y_{i1} and one attitudinal response r_{i1} are generated for each individual. An initial condition is not estimated in these specifications. Instead, variable y_{i0} replaced by a normally-distributed exogenous variable, drawn independently of λ_i. Thus, for cross-sectional specifications, the parameters do not suffer from an autocorrelation bias and the latent factor simply adds noise to the estimation. The same coefficient, sample size, latent-factor distribution, and factor loading options from the longitudinal specifications are used for the cross-sectional specifications.[6]

---

[6]This data generating process is summarized by the following steps:

1. Draw $y_{i0} \sim \mathcal{N}(0, 16)$, $X_{i1} \sim \mathcal{N}(0, 36)$, and λ_i from the specified distribution.
2. Construct latent indeces $y^*_{i1} = \beta^y_0 + \beta^y_1 y_{i0} + \beta^y_2 X_{i1} + \lambda_i$ and $r^*_{i1} = \beta^r_0 + \beta^r_2 X_{i1} + \rho^r \lambda_i$.
3. Draw $\Psi^y_{i1} \sim U[0, 1]$ and $\Psi^r_{i1} \sim U[0, 1]$.
4. Generate outcomes $y_{i1} = \mathbb{1}\left(\frac{\exp(y^*_{i1})}{1+\exp\left(y^*_{i1}\right)} \geq \Psi^y_{i1}\right)$ and $r_{i1} = \mathbb{1}\left(\frac{\exp(r^*_{i1})}{1+\exp\left(r^*_{i1}\right)} \geq \Psi^r_{i1}\right)$.

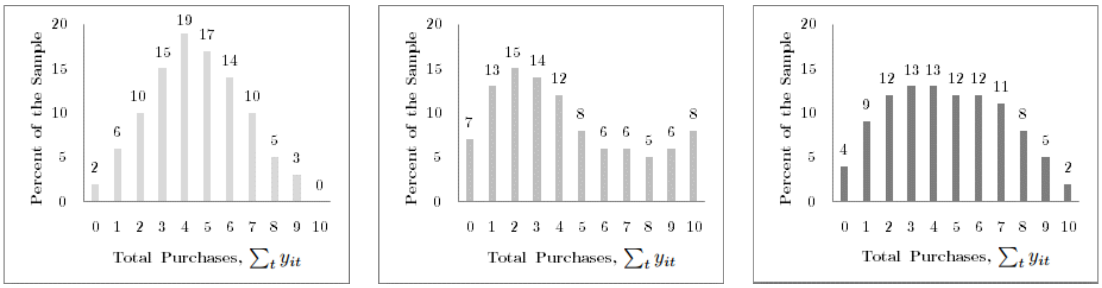
6.2.4 Monte Carlo Summary Statistics

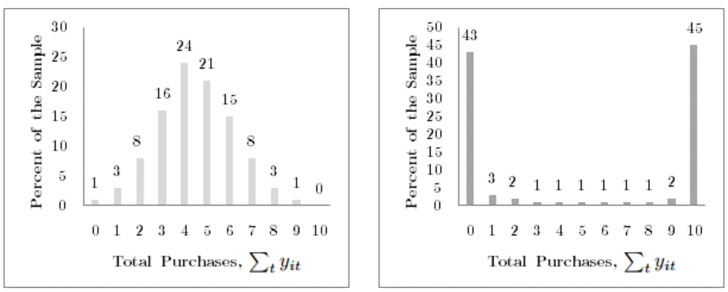
The two exercises in Section 6.2.4 quantify the impact of the latent factors on the data generating process (DGP) for outcomes y and r in the Monte Carlo experiment. For brevity, I analyze only specifications that include 1,000 individuals for 10 time periods. The exercises are repeated 2,000 times, and I average across these iterations to generate summary statistics. In order to isolate the impact of latent factors on the outcomes, only one complete set of exogenous covariates $\{X_{it}, W_{i0}\}$ and uniform random variables $\{\Psi_{it}^{y}, \Psi_{it}^{r}\}$ is drawn and used repeatedly throughout the exercises. Conditional on this fixed set of data, the only remaining variables are the latent factor (drawn at random from the specified distribution for every individual in each repetition) and the factor loading parameter ρ^r, which varies by specification.

The parameters for the distribution of latent factor λ_i were chosen such that the unobserved heterogeneity is *not* the dominant source of variation for outcome y. To verify this claim, the first exercise appeals to the time-invariant property of the latent factors. If, in the DGP, the impact of the time-invariant latent factors overpowers the influence of the time-varying exogenous characteristics, which is what I want to avoid, individuals will be very persistent in their purchasing behavior. Specifically, individuals who receive an extremely negative draw for latent factor λ_i will never purchase good y ($Y_i = \{0, 0, \ldots, 0\}$), and those who receive an extremely positive draw will always purchase the good ($Y_i = \{1, 1, \ldots, 1\}$). The frequency of total purchases ($\sum_t y_{it}$) in time periods $t = 1, \ldots, 10$ among the sample are displayed in Figure 6.2. As a baseline comparison, when I omit the latent factor from the DGP, good y is purchased zero times or ten times over the ten time periods by only 1% of the Monte Carlo sample. Under latent-factor Distribution 3, 2% of the sample purchases good y in each period and 4% never purchases the good. For Distribution 1, which has a lower variance, 2% of the sample purchases the good either zero or ten times. For Distribution 2, which has a higher variance, this percentage increases to 15%. Thus, the influence of time-variant exogenous characteristics in the DGP is not overshadowed by the permanent unobserved heterogeneity. Figure 6.2e shows an undesirable dispersion of frequencies that results from increasing the variance of Distribution 2 by a factor of 10. This dramatically increases the probability of receiving an extremely high or low draw for λ_i. The frequency of persistent purchasing habits across time jumps to 88%, suggesting that, in this extreme case, the individual's time-invariant latent-factor draw largely dictates the outcome realization.[7]

Figure 6.2: Population Frequencies: Total Purchases Per Individual ($\sum_t y_{it}$), by Latent-Factor Distribution

(a) Distribution 1 (b) Distribution 2 (c) Distribution 3





(d) Baseline, No Latent Factor (e) Extremely High Variance

An extremely high variance for the distribution of latent factor λ_i (Figure 6.2e) is undesirable because the time-invariant latent factor draw largely determines outcome y.

The factor loading parameter ρ^r plays a significant role in the DGP for outcome r. Ceteris paribus, latent factor λ_i has no impact on r when $\rho^r = 0$. However, a positive value for ρ^r generates a positive correlation between λ_i and r, similar to the positive correlation between λ_i and y. Thus, an increase in the magnitude of ρ^r strengthens the correlation in the error terms of outcomes y and r. For this second exercise, all other parameters and variables are held constant to quantify the impact of factor loading ρ^r on the generated response r. The off-diagonal elements of Contingency Table 6.2 report the percentages of outcomes that switch from 0 to 1 or from 1 to 0 following a change in the magnitude of ρ^r. The largest shift in outcome realizations occurs under Distribution 2 when the magnitude of ρ^r increases from zero to 0.9; 20% of the outcomes change after the increase.

[7] An identical conclusion results from repeating the exercise for outcome r under all specified values of ρ^r; the latent factor is not the dominant source of variation in the DGP for outcome r.

Table 6.2: Contingency Tables for Outcome r, by Magnitude of Factor Loading ρ^r

		r_t when $\rho^r = 0.3$			r_t when $\rho^r = 0.6$			r_t when $\rho^r = 0.9$		
			0	**1**		**0**	**1**		**0**	**1**
Latent-Factor Distribution 1	r_t when $\rho^r = 0$	**0**	60	2	**0**	58	3	**0**	57	5
		1	1	37	**1**	3	36	**1**	4	34
			0	**1**		**0**	**1**		**0**	**1**
Latent-Factor Distribution 2	r_t when $\rho^r = 0$	**0**	57	4	**0**	54	8	**0**	51	10
		1	4	35	**1**	7	31	**1**	10	29
			0	**1**		**0**	**1**		**0**	**1**
Latent-Factor Distribution 3	r_t when $\rho^r = 0$	**0**	59	3	**0**	56	6	**0**	53	8
		1	2	36	**1**	5	33	**1**	7	32

Off-diagonal elements represent the percent of outcomes affected by a change in the magnitude of ρ^r.

Summary statistics are reported in Tables 6.3 and 6.4. The penultimate row of Table 6.4 shows that the correlation across outcomes becomes more positive as factor loading ρ^r increases. The last row of Table 6.4 reports the correlation coefficient between the composite error terms in the DGP for y and r, which each consist of the latent factor and GEV shock.

Table 6.3: Monte Carlo Summary Statistics: $(N, T) = (1000, 10)$

		Mean	Std. Dev.	Minimum	Maximum
Covariates	X_t	-0.13	5.93	-22.48	21.55
	W_0	-0.20	6.00	-17.96	17.41

Table 6.4: Monte Carlo Summary Statistics: $(N, T) = (1000, 10)$

			Distribution 1				Distribution 2				Distribution 3			
			Factor Loading ρ^r				Factor Loading ρ^r				Factor Loading ρ^r			
			0	0.3	0.6	0.9	0	0.3	0.6	0.9	0	0.3	0.6	0.9
Composite Latent Factors	λ_Y	Mean	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		St.D.	1.08	1.08	1.08	1.08	2.94	2.94	2.94	2.94	1.83	1.83	1.83	1.83
		Min	-2.66	-2.80	-2.97	-2.90	-6.52	-6.50	-6.38	-6.46	-3.65	-4.07	-4.42	-4.25
		Max	2.99	3.07	3.17	3.13	11.80	11.77	11.87	11.77	4.02	4.40	4.77	4.59
	λ_R	Mean	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		St.D.	0.00	0.33	0.65	0.97	0.00	0.88	1.76	2.65	0.00	0.55	1.10	1.65
		Min	0.00	-0.84	-1.78	-2.61	0.00	-1.95	-3.83	-5.81	0.00	-1.22	-2.65	-3.83
		Max	0.00	0.92	1.90	2.82	0.00	3.53	7.12	10.59	0.00	1.32	2.86	4.13
Outcomes	y_t	Mean	0.43	0.43	0.43	0.43	0.42	0.42	0.42	0.42	0.44	0.44	0.44	0.44
	r_t	Mean	0.38	0.39	0.39	0.39	0.39	0.39	0.40	0.39	0.38	0.38	0.39	0.40
		Corr.	-0.44	-0.42	-0.39	-0.37	-0.33	-0.20	-0.07	0.03	-0.41	-0.34	-0.27	-0.20
Corr. across composite error			0.00	0.15	0.28	0.38	0.00	0.52	0.75	0.82	0.00	0.32	0.53	0.64

Here, $\lambda_Y = \lambda_i$ and $\lambda_R = \rho^r \lambda_i$. The composite error terms are $\{\lambda_Y$ plus a GEV shock$\}$ and $\{\lambda_R$ plus a GEV shock$\}$. Data for each specification are simulated 1,000 times. Averages are reported.

6.2.5 Statistical Models Compared in the Estimation of the Monte Carlo Experiment

To test the econometric theory in Section 6.1, I compare a Discrete-Factor Random Effects (DFRE) model that does not use attitudinal data (**Model I**) to a DFRE model that jointly estimates both observed outcomes y and r (**Model C**, *the preferred model*). I also compare the preferred model against other empirical methods found in the literature, including a standard logit model (**Model L1**), a logit model with attitudinal responses as explanatory variables (**Model L2**), and a DFRE model with attitudinal responses as explanatory variables (**Model E**). The Monte Carlo experiment simulates 500 repetitions for each of the 96 specifications. The latent factors used to generate the data are unobserved by the econometrician. The statistical models for longitudinal specifications are outlined below. For cross-sectional specifications, the initial condition equation in each model is omitted because only time period $t = 1$ is estimated.

1. **Model L1** – Logit without Attitudinal Responses

$$\begin{aligned} \ln\left(\tfrac{\Pr(y_{i0}=1)}{\Pr(y_{i0}=0)}\right) &= \beta_0^{y_0} + \beta_1^{y_0} W_{i0} \\ \ln\left(\tfrac{\Pr(y_{it}=1)}{\Pr(y_{it}=0)}\right) &= \beta_0^{y} + \beta_1^{y} y_{it-1} + \beta_2^{y} X_{it} \quad \text{for } t = 1, ..., T-1 \end{aligned} \tag{6.4}$$

2. **Model L2** – Logit with Attitudinal Responses as Explanatory Variables

$$\begin{aligned} \ln\left(\tfrac{\Pr(y_{i0}=1)}{\Pr(y_{i0}=0)}\right) &= \beta_0^{y_0} + \beta_1^{y_0} W_{i0} \\ \ln\left(\tfrac{\Pr(y_{it}=1)}{\Pr(y_{it}=0)}\right) &= \beta_0^{y} + \beta_1^{y} y_{it-1} + \beta_2^{y} X_{it} + \beta_3^{y} r_{it} \quad \text{for } t = 1, ..., T-1 \end{aligned} \tag{6.5}$$

3. **Model I** – DFRE, Logit

$$\begin{cases} \ln\left(\tfrac{\Pr(y_{i0}=1)}{\Pr(y_{i0}=0)}\right) = \beta_0^{y_0} + \beta_1^{y_0} W_{i0} + \lambda^{y_0} \\ \ln\left(\tfrac{\Pr(y_{it}=1)}{\Pr(y_{it}=0)}\right) = \beta_0^{y} + \beta_1^{y} y_{it-1} + \beta_2^{y} X_{it} + \lambda^{y} \quad \text{for } t = 1, ..., T-1 \end{cases} \tag{6.6}$$

4. **Model E** – DFRE, Logit with Attitudinal Responses as Explanatory Variables

$$\begin{cases} \ln\left(\frac{\Pr(y_{i0}=1)}{\Pr(y_{i0}=0)}\right) = \beta_0^{y_0} + \beta_1^{y_0} W_{i0} + \lambda^{y_0} \\ \ln\left(\frac{\Pr(y_{it}=1)}{\Pr(y_{it}=0)}\right) = \beta_0^{y} + \beta_1^{y} y_{it-1} + \beta_2^{y} X_{it} + \beta_3^{y} r_{it} + \lambda^{y} \quad \text{for } t = 1, ..., T-1 \end{cases} \tag{6.7}$$

5. **Model C** – DFRE, Logit of Jointly-Estimated Outcomes y and Attitudinal Responses r

$$\begin{cases} \ln\left(\frac{\Pr(y_{i0}=1)}{\Pr(y_{i0}=0)}\right) = \beta_0^{y_0} + \beta_1^{y_0} W_{i0} + \lambda^{y_0} \\ \ln\left(\frac{\Pr(y_{it}=1)}{\Pr(y_{it}=0)}\right) = \beta_0^{y} + \beta_1^{y} y_{it-1} + \beta_2^{y} X_{it} + \lambda^{y} \quad \text{for } t = 1, ..., T-1 \\ \ln\left(\frac{\Pr(r_{it}=1)}{\Pr(r_{it}=0)}\right) = \beta_0^{r} + \beta_1^{r} y_{it-1} + \beta_2^{r} X_{it} + \lambda^{r} \quad \text{for } t = 1, ..., T-1 \end{cases} \tag{6.8}$$

Chapter 5 explained the DFRE estimation routine in detail. Briefly, taking **Model C** as an example, the DFRE method estimates several mass points of the form $\widehat{\lambda_k} = \left\{\widehat{\lambda_k^{y_0}}, \widehat{\lambda_k^{y}}, \widehat{\lambda_k^{r}}\right\}$ and assigns a probability to each mass point, in which $\widehat{\mathfrak{p}_k} = \Pr\left(\lambda = \widehat{\lambda_k}\right)$. These parameters are used in the likelihood function to integrate over the distribution of the permanent unobserved heterogeneity. The Bayesian Information Criterion (BIC) is used to select the optimal number of mass points in each repetition.

6.2.6 Hypotheses

In longitudinal specifications, explanatory variable y_{it-1} and dependent variable y_{it} are both functions of latent factor λ_i. **Models L1** and **L2** do not address the serial autocorrelation, which should result in biased estimates for coefficient $\widehat{\beta_1^y}$. **Models I**, **E**, and **C** incorporate random effects to address the serial autocorrelation by estimating the distribution of the unobserved heterogeneity, which improves the accuracy of the estimated coefficients. For cross-sectional specifications, no estimated coefficients suffer from an endogeneity bias, but the latent factor λ_i still introduces a significant amount of noise to the model. I examine the case in which the econometrician is only interested in the marginal effects of explanatory variables X_{it} and y_{it-1} on the demand for y_{it}. From these considerations, I form the following hypotheses about the estimation of β_1^y and β_2^y:

1. For any specification, **Models L1** and **L2** should produce the least accurate estimates because these models do not control for unobserved heterogeneity.

2. For specifications in which $\rho^r = 0$, there is no correlation among the error terms, and **Model I** should produce the most accurate and efficient estimates. The incorporation of attitudinal variables in **Models E** and **C** only adds noise to the model.

3. For specifications in which $\rho^r > 0$, **Model C** should produce more efficient point estimates than **Model I**. The comparison of small-sample accuracy between **Models I** and **C** remains an experimental question, as the econometric theory only suggests that both estimators are asymptotically unbiased.

4. The comparison between **Models E** and **C** also remains an experimental question; the econometric theory did not offer a direct comparison between these models.

6.2.7 Monte Carlo Simulation Results: Accuracy

McCullagh and Nelder (1989) show that, as $N \to \infty$, the parameter point estimates do not differ between **Models I** and **C**. But, does the inclusion of attitudinal outcomes reduce the bias for small sample sizes? After running 500 simulations for a particular specification, I compute the Mean Absolute Deviation (MAD) between the estimated and true coefficients to gauge each model's accuracy. A model that produces the lowest MAD is considered most accurate because this model estimates a parameter closest to the true coefficient, on average. Tables 6.5−6.6, 6.7−6.8, and 6.9−6.10 summarize the MAD for each model when latent factors are drawn from Distribution 1, 2, and 3, respectively.

For the cross-sectional Distribution 1 specifications (the first four rows of Tables 6.5 and 6.6), the parameters do not suffer from an endogeneity bias, the sample size is small, and the variance of latent-factor distribution is low. The standard logit models (**L1** and **L2**) produce the lowest MAD for seven of these eight specifications. However, when serial autocorrelation is introduced into the model (rows five through twelve), the standard logit models typically produce the least accurate estimates. Furthermore, when the variance of the latent-factor distribution increases (Distributions 2 and 3 in Tables 6.7−6.10), **Models L1** and **L2** perform significantly worse than the DFRE models.

For most specifications in which $\rho^r = 0$, **Model C** is the least accurate. Jointly estimating outcomes y and r offers no benefit to estimation if the error terms are statistically independent. However, the MAD of the estimated coefficients from **Model C** shrinks as the magnitude of factor loading ρ^r increases. This result suggests that the accuracy of the response parameters for **Model C** in small samples improves as the correlation between error terms strengthens. When the latent factor is drawn from Distribution 1 and its factor loading is $\rho^r = 0.9$, **Model C** achieves a lower MAD than **Models I** and **E** for both β_1^Y and β_2^y in all but one specification. For specifications in which the latent factor is drawn from Distribution 2, **Model C** consistently achieves a lower MAD than **Model I** when $\rho^r = 0.9$ and achieves the lowest MAD in all but two specifications. When $\rho^r = 0.9$ and λ_i is drawn from Distribution 3, **Model C** achieves the lowest MAD for β_2^y in all but one specification and the lowest MAD for β_1^y in half of the specifications. The econometric theory concludes that both models will produce asymptotically unbiased coefficient estimates regardless of the strength in correlation between error terms. But, this Monte Carlo experiment suggests that jointly estimating both outcomes actually improves the accuracy of the point estimates in small samples when this correlation is strong.

In the following tables, the lowest MAD of the DFRE models is **boldfaced** for each specification, uniquely identified by the latent-factor distribution, sample size N, time periods T, and magnitude parameter ρ^r. For the few instances in which **Model L1** or **L2** achieves the lowest MAD, these numbers are <u>underlined</u>.

Table 6.5: Mean Absolute Deviation from the True β_1^y (Response Parameter for y_{t-1}), Distribution 1

(N,T)	ρ^r	**L1**	**L2**	**I**	**E**	**C**
(500, 1)	0.0	.106	<u>.104</u>	.162	.168	**.155**
	0.3	.104	<u>.101</u>	.175	.181	**.161**
	0.6	.107	<u>.102</u>	.163	.169	**.159**
	0.9	.098	<u>.091</u>	.173	.178	**.161**
(500, 3)	0.0	.327	.328	**.203**	.205	.224
	0.3	.327	.325	**.203**	.204	.211
	0.6	.327	.315	.203	.202	**.198**
	0.9	.320	.296	.194	.193	**.187**
(500, 5)	0.0	.329	.330	.138	**.138**	.145
	0.3	.340	.337	**.125**	.127	.126
	0.6	.325	.311	.135	.134	**.133**
	0.9	.330	.300	.131	.131	**.128**
(500, 10)	0.0	.348	.348	**.088**	.088	.088
	0.3	.346	.342	**.087**	.087	.087
	0.6	.341	.325	**.086**	.087	.088
	0.9	.347	.315	.086	.088	**.086**

Average across 500 repetitions for each specification

(N,T)	ρ^r	**L1**	**L2**	**I**	**E**	**C**
(1000, 1)	0.0	.103	<u>.102</u>	**.104**	.108	.109
	0.3	.103	<u>.101</u>	**.123**	.123	.123
	0.6	.106	<u>.102</u>	.104	.105	**.102**
	0.9	.109	.101	.106	.107	**.098**
(1000, 3)	0.0	.317	.318	**.136**	.137	.158
	0.3	.315	.312	.140	**.139**	.143
	0.6	.323	.309	.134	.133	**.132**
	0.9	.326	.299	.138	.138	**.137**
(1000, 5)	0.0	.341	.342	**.089**	.090	.089
	0.3	.334	.330	.097	**.096**	.097
	0.6	.326	.311	.089	.091	**.089**
	0.9	.331	.301	**.086**	.089	.091
(1000, 10)	0.0	.348	.348	**.067**	.067	.070
	0.3	.343	.338	**.065**	.065	.067
	0.6	.347	.331	.067	.069	**.066**
	0.9	.342	.310	.063	.066	**.056**

Average across 500 repetitions for each specification

Table 6.6: Mean Absolute Deviation from the True β_2^y (Response Parameter for X_t), Distribution 1

(N,T)	ρ^r	**L1**	**L2**	**I**	**E**	**C**
(500, 1)	0.0	<u>.070</u>	.071	.107	.111	**.103**
	0.3	.070	<u>.058</u>	.114	.130	**.106**
	0.6	.070	<u>.052</u>	.112	.139	**.109**
	0.9	.067	<u>.044</u>	.113	.152	**.108**
(500, 3)	0.0	.071	.070	**.041**	.044	.048
	0.3	.071	.052	**.041**	.047	.043
	0.6	.071	<u>.038</u>	.041	.052	**.039**
	0.9	.073	<u>.031</u>	.043	.058	**.037**
(500, 5)	0.0	.071	.069	**.021**	.024	.023
	0.3	.072	.055	**.024**	.027	.024
	0.6	.071	.037	.022	.029	**.022**
	0.9	.072	.026	.023	.035	**.022**
(500, 10)	0.0	.072	.071	**.016**	.017	.016
	0.3	.073	.054	.017	**.016**	.017
	0.6	.072	.037	**.015**	.017	.016
	0.9	.072	.023	.016	.022	**.015**

Average across 500 repetitions for each specification

(N,T)	ρ^r	**L1**	**L2**	**I**	**E**	**C**
(1000, 1)	0.0	<u>.068</u>	.069	**.070**	.076	.073
	0.3	.069	<u>.055</u>	**.082**	.091	.083
	0.6	.071	<u>.044</u>	.068	.087	**.066**
	0.9	.073	<u>.036</u>	.069	.095	**.066**
(1000, 3)	0.0	.073	.072	**.028**	.030	.034
	0.3	.073	.054	**.028**	.031	.029
	0.6	.071	.036	.028	.037	**.026**
	0.9	.072	.025	.027	.042	**.023**
(1000, 5)	0.0	.072	.071	**.015**	.018	.015
	0.3	.073	.054	.016	.017	**.015**
	0.6	.073	.037	**.015**	.023	.016
	0.9	.073	.023	.016	.031	**.016**
(1000, 10)	0.0	.073	.072	**.012**	.013	.012
	0.3	.073	.055	.012	**.011**	.013
	0.6	.073	.037	.012	.014	**.011**
	0.9	.073	.023	.012	.020	**.010**

Average across 500 repetitions for each specification

Table 6.7: Mean Absolute Deviation from the True β_1^y (Response Parameter for y_{t-1}), Distribution 2

(N,T)	ρ^r	**L1**	**L2**	**I**	**E**	**C**
$(500,1)$	0.0	.374	.373	**.244**	.259	.275
	0.3	.375	.359	.246	.248	**.188**
	0.6	.373	.327	.242	.211	**.144**
	0.9	.373	.297	.236	.189	**.131**
$(500,3)$	0.0	1.280	1.284	.351	**.351**	.356
	0.3	1.280	1.221	**.351**	.386	.369
	0.6	1.280	1.086	**.351**	.418	.401
	0.9	1.275	.952	**.333**	.395	.348
$(500,5)$	0.0	1.373	1.374	.205	**.205**	.243
	0.3	1.354	1.282	.204	**.204**	.205
	0.6	1.373	1.153	.214	.234	**.188**
	0.9	1.359	1.010	.202	.273	**.191**
$(500,10)$	0.0	1.427	1.427	**.119**	.120	.126
	0.3	1.423	1.343	**.116**	.116	.119
	0.6	1.433	1.201	.123	.126	**.116**
	0.9	1.427	1.057	.121	.133	**.113**

Average across 500 repetitions for each specification

(N,T)	ρ^r	**L1**	**L2**	**I**	**E**	**C**
$(1000,1)$	0.0	.377	.376	**.233**	.245	.262
	0.3	.378	.364	.236	.266	**.174**
	0.6	.375	.331	.241	.234	**.154**
	0.9	.375	.301	.239	.204	**.133**
$(1000,3)$	0.0	1.275	1.277	**.266**	.269	.293
	0.3	1.294	1.232	**.291**	.312	.316
	0.6	1.300	1.101	.292	.362	**.261**
	0.9	1.295	.971	.289	.388	**.204**
$(1000,5)$	0.0	1.366	1.366	**.121**	.121	.124
	0.3	1.365	1.291	**.126**	.130	.132
	0.6	1.365	1.146	**.129**	.156	.133
	0.9	1.364	1.014	.131	.190	**.123**
$(1000,10)$	0.0	1.422	1.422	**.079**	.079	.080
	0.3	1.423	1.341	**.076**	.077	.079
	0.6	1.422	1.192	**.079**	.081	.083
	0.9	1.426	1.057	.075	.086	**.074**

Average across 500 repetitions for each specification

Table 6.8: Mean Absolute Deviation from the True β_2^y (Response Parameter for X_t), Distribution 2

(N,T)	ρ^r	**L1**	**L2**	**I**	**E**	**C**
	0.0	.250	.247	**.164**	.175	.184
(500, 1)	0.3	.249	.168	.164	**.118**	.123
	0.6	.249	.112	.162	.108	**.093**
	0.9	.248	.086	.157	.112	**.078**
	0.0	.221	.211	**.073**	.075	.073
(500, 3)	0.3	.221	.145	.073	**.050**	.076
	0.6	.221	.102	.073	**.042**	.072
	0.9	.224	.085	.075	**.035**	.060
	0.0	.220	.211	**.039**	.040	.050
(500, 5)	0.3	.219	.146	.040	**.033**	.038
	0.6	.219	.104	.041	.038	**.033**
	0.9	.219	.082	.037	.039	**.032**
	0.0	.218	.209	**.022**	.023	.026
(500, 10)	0.3	.218	.145	.022	**.019**	.021
	0.6	.218	.104	.022	**.020**	.021
	0.9	.218	.084	.022	.020	**.018**

Average across 500 repetitions for each specification

(N,T)	ρ^r	**L1**	**L2**	**I**	**E**	**C**
	0.0	.249	.246	**.154**	.160	.173
(1000, 1)	0.3	.252	.172	.157	**.112**	.114
	0.6	.249	.117	.160	**.083**	.097
	0.9	.251	.089	.159	**.072**	.080
	0.0	.222	.214	**.061**	.065	.071
(1000, 3)	0.3	.222	.146	.060	**.045**	.065
	0.6	.221	.102	.063	**.038**	.045
	0.9	.222	.082	.060	**.030**	.032
	0.0	.220	.211	**.026**	.029	.027
(1000, 5)	0.3	.220	.146	.026	**.020**	.027
	0.6	.220	.105	.025	.028	**.025**
	0.9	.220	.085	.025	.031	**.020**
	0.0	.218	.208	**.013**	.013	.013
(1000, 10)	0.3	.218	.146	**.012**	.013	.013
	0.6	.218	.105	**.013**	.017	.013
	0.9	.218	.084	.012	.022	**.011**

Average across 500 repetitions for each specification

Table 6.9: Mean Absolute Deviation from the True β_1^y (Response Parameter for y_{t-1}), Distribution 3

(N,T)	ρ^r	**L1**	**L2**	**I**	**E**	**C**
(500, 1)	0.0	.236	.234	**.137**	.142	.176
	0.3	.236	.230	**.137**	.140	.154
	0.6	.237	.218	.138	.145	**.137**
	0.9	.234	.198	.137	.150	**.109**
(500, 3)	0.0	.813	.815	.235	.237	**.235**
	0.3	.813	.795	.235	.240	**.233**
	0.6	.813	.744	.235	.244	**.234**
	0.9	.814	.682	**.253**	.258	.262
(500, 5)	0.0	.820	.821	**.176**	.177	.178
	0.3	.821	.798	**.168**	.171	.174
	0.6	.821	.744	**.168**	.179	.183
	0.9	.818	.675	**.172**	.191	.179
(500, 10)	0.0	.841	.841	.091	**.091**	.099
	0.3	.841	.817	**.090**	.091	.092
	0.6	.845	.763	**.094**	.098	.095
	0.9	.838	.689	**.091**	.100	.091

Average across 500 repetitions for each specification

(N,T)	ρ^r	**L1**	**L2**	**I**	**E**	**C**
(1000, 1)	0.0	.237	.236	**.100**	.100	.154
	0.3	.241	.235	**.105**	.112	.128
	0.6	.240	.221	.102	.103	**.095**
	0.9	.239	.203	.100	.107	**.079**
(1000, 3)	0.0	.802	.803	**.173**	.174	.175
	0.3	.810	.791	**.189**	.195	.193
	0.6	.805	.733	**.181**	.194	.191
	0.9	.799	.664	**.176**	.195	.196
(1000, 5)	0.0	.830	.830	**.141**	.142	.149
	0.3	.819	.796	**.126**	.131	.133
	0.6	.824	.748	.138	.152	**.129**
	0.9	.820	.677	.132	.160	**.104**
(1000, 10)	0.0	.848	.848	**.070**	.070	.070
	0.3	.843	.817	**.067**	.067	.067
	0.6	.847	.765	**.069**	.071	.071
	0.9	.850	.699	.068	.075	**.064**

Average across 500 repetitions for each specification

Table 6.10: Mean Absolute Deviation from the True β_2^y (Response Parameter for X_t), Distribution 3

(N,T)	ρ^r	**L1**	**L2**	**I**	**E**	**C**
	0.0	.160	.159	**.091**	.097	.119
(500, 1)	0.3	.160	.119	**.091**	.108	.105
	0.6	.157	$\underline{.078}$	**.090**	.125	.091
	0.9	.156	$\underline{.056}$	.089	.136	**.072**
	0.0	.151	.150	**.040**	.043	.040
(500, 3)	0.3	.151	.109	.040	.041	**.039**
	0.6	.151	.073	.040	.048	**.038**
	0.9	.151	.049	.042	.053	**.040**
	0.0	.151	.149	**.033**	.034	.033
(500, 5)	0.3	.152	.111	.034	**.026**	.035
	0.6	.152	.076	.034	**.028**	.035
	0.9	.153	.051	.034	.032	**.029**
	0.0	.150	.149	**.015**	.016	.016
(500, 10)	0.3	.151	.111	**.015**	.017	.016
	0.6	.151	.075	.016	.024	**.016**
	0.9	.151	.051	**.015**	.028	.016

Average across 500 repetitions for each specification

(N,T)	ρ^r	**L1**	**L2**	**I**	**E**	**C**
	0.0	.159	.158	**.069**	.070	.105
(1000, 1)	0.3	.161	.120	**.068**	.074	.085
	0.6	.159	.079	.069	.084	**.064**
	0.9	.160	.053	.067	.101	**.052**
	0.0	.153	.152	**.036**	.037	.036
(1000, 3)	0.3	.153	.111	.034	**.027**	.035
	0.6	.152	.075	**.034**	.035	.035
	0.9	.153	.050	.036	.042	**.035**
	0.0	.152	.151	**.028**	.029	.031
(1000, 5)	0.3	.152	.111	.028	**.019**	.031
	0.6	.152	.077	.030	**.023**	.024
	0.9	.152	.051	.030	.030	**.017**
	0.0	.152	.151	.011	.012	**.011**
(1000, 10)	0.3	.152	.111	**.011**	.012	.012
	0.6	.152	.077	**.011**	.019	.012
	0.9	.152	.051	.011	.029	**.010**

Average across 500 repetitions for each specification

6.2.8 Monte Carlo Simulation Results: Efficiency

For Monte Carlo experiments, efficiency is measured by the root mean squared error (RMSE). Because **Models L1** and **L2** were generally the least accurate, I omit **Models L1** and **L2** from the following analysis. Tables 6.11 and 6.12 summarize both the MAD and efficiency gain comparisons between **Model C** and **Model I**. If **Model C** is more accurate (lower MAD) than **Model I**, the relative efficiency gain is reported for this specification. Otherwise, a hyphen appears in the cell. Consistent with the econometric theory, **Model C** does provide efficiency gains over **Model I** when there is correlation between the error terms ($\rho^r > 0$), and the efficiency generally improves as the correlation increases. The relative efficiency gain formula, in which β is the true coefficient value and $\widehat{\beta}_s$ is the coefficient estimate from repetition s, is given by:

$$\text{Relative Efficiency Gain} = \frac{\text{RMSE}^{M(I)} - \text{RMSE}^{M(C)}}{\text{RMSE}^{M(I)}}, \text{ where RMSE} = \sqrt{\sum_s \left(\widehat{\beta}_s - \beta\right)^2}. \quad (6.9)$$

Table 6.11: Coefficient β_1^y Efficiency Gains for
Model C Relative to **Model I**, by Specification

	Distribution 1				Distribution 2				Distribution 3			
	Factor Loading ρ^r				Factor Loading ρ^r				Factor Loading ρ^r			
(N,T)	0	0.3	0.6	0.9	0	0.3	0.6	0.9	0	0.3	0.6	0.9
$(500,1)$	9	11	3	7	–	40	64	69	–	–	5	32
$(500,3)$	–	–	7	8	–	–	–	–	-1	1	0	–
$(500,5)$	–	–	5	4	–	–	22	10	–	–	–	–
$(500,10)$	–	–	–	1	–	–	8	11	–	–	–	–
$(1000,1)$	–	–	5	12	–	49	59	69	–	–	15	38
$(1000,3)$	–	–	3	3	–	–	13	42	–	–	–	–
$(1000,5)$	–	–	1	–	–	–	–	10	–	–	10	36
$(1000,10)$	–	–	2	17	–	–	–	2	–	–	–	13

The relative RMSE is reported if **Model C** achieves a lower MAD than **Model I**.

Table 6.12: Coefficient β_2^y Efficiency Gains for
Model C Relative to **Model I**, by Specification

	Distribution 1				Distribution 2				Distribution 3			
	Factor Loading ρ^r				Factor Loading ρ^r				Factor Loading ρ^r			
(N, T)	0	0.3	0.6	0.9	0	0.3	0.6	0.9	0	0.3	0.6	0.9
$(500, 1)$	7	10	4	6	–	43	67	74	–	–	–	32
$(500, 3)$	–	–	11	22	–	–	3	30	–	2	9	14
$(500, 5)$	–	–	4	10	–	8	37	33	–	–	–	24
$(500, 10)$	–	1	–	4	–	0	13	30	–	–	3	–
$(1000, 1)$	–	–	4	8	–	50	63	75	–	–	17	39
$(1000, 3)$	–	–	16	32	–	–	40	67	–	–	–	3
$(1000, 5)$	–	1	–	0	–	–	3	34	–	–	33	63
$(1000, 10)$	–	–	18	35	–	–	–	18	0	–	–	18

The relative RMSE is reported if **Model C** achieves a lower MAD than **Model I**.

Tables 6.13 and 6.14 show the RMSE comparison between **Model C** and **Model E**, in which the attitudinal response r is treated as an exogenous variable in the equation for outcome y. The results are similar to Tables 6.11 and 6.12; as the correlation between error terms strengthens, the efficiency gains for **Model C** over **Model E** increases.

Table 6.13: Coefficient β_1^y Efficiency Gains for
Model C Relative to **Model E**, by Specification

	Distribution 1				Distribution 2				Distribution 3			
	Factor Loading ρ^r				Factor Loading ρ^r				Factor Loading ρ^r			
(N, T)	0	0.3	0.6	0.9	0	0.3	0.6	0.9	0	0.3	0.6	0.9
$(500, 1)$	15	19	11	16	–	38	53	51	–	–	11	40
$(500, 3)$	–	–	5	7	–	7	7	16	1	5	6	–
$(500, 5)$	–	0	2	4	–	–	33	47	–	–	–	15
$(500, 10)$	–	0	–	5	–	–	13	24	–	–	8	18
$(1000, 1)$	–	–	8	15	–	57	52	56	–	–	16	45
$(1000, 3)$	–	–	2	3	–	–	38	63	–	1	1	–
$(1000, 5)$	0	–	4	–	–	–	23	55	–	–	24	53
$(1000, 10)$	–	–	8	24	–	–	–	26	1	0	–	26

The relative RMSE is reported if **Model C** achieves a lower MAD than **Model E**.

Table 6.14: Coefficient β_2^y Efficiency Gains for
Model C Relative to **Model E**, by Specification

	Distribution 1				Distribution 2				Distribution 3			
	Factor Loading ρ^r				Factor Loading ρ^r				Factor Loading ρ^r			
(N, T)	0	0.3	0.6	0.9	0	0.3	0.6	0.9	0	0.3	0.6	0.9
$(500, 1)$	13	34	37	45	–	–	41	60	–	17	52	75
$(500, 3)$	–	20	48	59	10	–	–	–	11	9	41	49
$(500, 5)$	-3	16	46	60	–	–	29	38	9	–	–	24
$(500, 10)$	19	–	17	53	–	–	–	22	0	21	56	67
$(1000, 1)$	7	23	42	54	–	–	–	–	–	–	48	76
$(1000, 3)$	–	18	53	71	–	–	–	–	9	–	–	33
$(1000, 5)$	22	25	51	72	13	–	20	56	–	–	–	69
$(1000, 10)$	9	–	42	72	-2	–	33	71	16	–	58	84

The relative RMSE is reported if **Model C** achieves a lower MAD than **Model E**.

Chapter 7

Data Application

Starting with an initial sample of 2,379 nine to ten year old girls in 1987, the National Heart, Lung, and Blood Institute Growth and Health Study (NGHS) surveyed these girls once a year through age 18 or 19. The study was conducted in Washington D.C., Cincinnati (OH), and Richmond (CA) and oversampled black girls to achieve an even 50-50 white-to-black ratio for the sample. Surveyors in Cincinnati and Richmond took applications from children in public and parochial schools from both urban and suburban areas to generate variation in socioeconomic attributes. For D.C., only Group Health Association HMO subscribers were considered. Applications were approved if there was at least one parent willing to submit family demographic information, if the family had no current plans to move from the area, and if the girl had nearly 100% black or white heritage. Results from the child's detailed annual physical examination served as the incentive for participating in the program, and parents were also reimbursed for any transportation costs associated with traveling to and from the research lab. All surveys were administered at the research facility without parental supervision. Participants were reminded in several parts of the survey that their answers would be kept confidential and would "not be shared with parents, teachers, or friends."

7.1 Study Objective and Contents

The NGHS primarily sought to determine if eating habits, activity patterns, socioeconomic factors, or psychosocial characteristics could explain differences between black and white girls in the prevalence of obesity and heart disease risk factors. The contents of the survey include:

1. Demographics: Participants provided information on their household demographics.

2. Complete Physical Examination: Clinicians took detailed measurements of the child's height, weight, body circumferences, and physical maturation stage. They also recorded the results from a blood count analysis.

3. Nutrition Diary: Nutritionists first showed participants how to measure their food and record entries in the diary. The children (or parents) were responsible for filling out the food diaries, which were maintained for three consecutive days. Each distinct food item is listed separately and organized according to its general category. Nutritional values for each food item are based off of the category's typical nutritional facts. For example, if the girl drank and wrote down "*8 fl.oz. of Hawaiian Punch*", the data would record 261 GRAMS of "FRUIT FLAVORED DRINK" with 131 calories, 3.4 g protein, 31 g carbohydrates, etc.

4. Physical Activity Diary: Participants kept track of their everyday activities in the physical activity diary by recording each activity, such as "BICYCLING," and its duration. Survey administrators then translated the entries into a physical activity score, based on the metabolic equivalent of task (MET) ranking for significant activities throughout the day. For example, 15-30 MINUTES of "SITTING WITH TV OR BOOK" would receive a lower score than 15-30 MINUTES of "RUNNING."

5. Nutritional Patterns: The survey asked participants about their typical drug, alcohol, and contraceptive usages. Other parts of this survey asked questions such as "How often do you eat fast food?", "Do you eat breakfast?", "How often do you snack?", "Do you read nutrition labels?", and "Who cooks dinner at your house?"

6. Psychological Assessment: This section asked self-perception questions about the child's personality, health beliefs, opinions of her social environment, and behavior when dealing with problems.

7. Parent Survey: A similar self-reported survey (consisting of parts 1, 2, 5, and 6 from the child's survey) was distributed to parents in odd numbered grade years, to be mailed back after completion. The response rate after the first year was very low.

7.2 Available Data

Due to funding restrictions, some of the survey forms were not distributed in all ten years. Data on the child's demographics, physical examination, and health behaviors were collected in every year of the study. Several forms contained attitudinal questions, but the study alternated among these forms each year. Other forms were administered only a few times. Table 7.1 summarizes the available data and indicates where the outcome variables, attitudinal responses, and exogenous characteristics are found. The checkmark signifies that the form was distributed in this year and that a significant number of participants completed the questions. The attitudinal questions in this data application are taken from categories Self-Esteem and "My Problems" and "How I Deal With Things."

Table 7.1: Available Data, by Grade Year

Type	Survey Form Grade Year →	3	4	5	6	7	8	9	10	11	12
X	Main Information, Demographics	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
X	Life Events								✓		✓
X	Physical Exam	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
X	Blood Tests	✓		✓		✓		✓			✓
X	Parent Survey	✓									
Y	Smoking and Drinking History	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Y	Food and Exercise Diaries	✓	✓	✓	✓	✓		✓	✓		✓
Y	Nutritional Habits	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Y	Physical Activity Patterns	✓		✓		✓	✓	✓	✓	✓	✓
R	Health Beliefs	✓		✓		✓		✓		✓	
R	"My Feelings"			✓		✓		✓		✓	✓
R	Self-Esteem and "My Problems"		✓		✓		✓		✓		✓
R	"How I Deal With Things"		✓		✓		✓		✓		✓
R	"What I Think and Feel"			✓		✓					
R	"What I Am Like"	✓		✓		✓					

Legend: X → Exogenous Variables; Y → Outcome Variables; R → Attitudinal Response Variables; ✓: Available; Empty: Unavailable

For the estimation sample and summary statistics presented in this section, an individual's information from a particular year is included only if the individual responds to the smoking and attitudinal questions included in the data application for that year and every previous year. The dynamic transition statistics are calculated using only individuals who remain in this sample for two consecutive years. Table 7.2 shows how the sample size changes throughout the survey period. In total, 18,000 of the possible 23,790 person-year observations report all dependent variables.[1] The NGHS lost a significant number of participants after the fifth wave

[1] 1,092 individuals eventually exit the survey, based on my requirements. In the first year that they

because the original study was approved for only five years and some parents opted out after the initial contract expired.

Table 7.2: Data Application Sample Size, by Grade Year

Grade Year →	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Observations	2,379	2,247	2,160	2,047	1,913	1,648	1,536	1,441	1,341	1,288
Percent Lost	–	6	4	5	7	14	7	6	7	4
Percent Remaining	100	94	91	86	80	69	65	61	56	54

Total observations = 18,000.

7.3 Smoking Habits: Data and Summary Statistics

Data describing the individual's smoking habits provide the best application for the empirical model in Chapter 5; the unobserved factors influencing these outcomes might also affect the individual's attitudinal responses, the data are available in every year of the survey, and the individual can easily quantify her smoking frequency by marking the appropriate category. In the first five years of the study, for grades 3 to 7, participants are asked if they have smoked cigarettes in the last year. For the remaining survey years, for grades 8 to 12, the child provides her frequency of use in terms of days per month. These questions, as they appear on the survey, are given below:

Grade Year	Question and Response Categories	
3 – 7	**Have you smoked any cigarettes in the past year?**	
	Responses:	Yes, No
8 – 12	**During the past 30 days, on how many days did you smoke cigarettes?**	
	Responses:	I did not smoke cigarettes during the past 30 days,
		1 or 2 days, 3-5, 6-10, 11-19, 20-29, All 30 days

Table 7.3 displays the summary statistics for these variables by year. In the first year, no one reports any cigarette use, but this percentage steadily increases each year. In the final survey

leave, 924 do not answer any of the survey, 105 do not provide smoking information, and 63 do not provide attitudinal responses. The remaining 4,698 omitted observations are excluded because the individual had previously exited the survey.

period, 30% of the sample report some smoking. For the data application, I will simplify the smoking variable to a binary outcome that separates the smokers from the non-smokers.

Table 7.3: Reported Smoking Habits, Percent of Sample by Grade

	Grade Year:						**Grade Year:**				
Current Smoker	**3**	**4**	**5**	**6**	**7**	Smoking Habits	**8**	**9**	**10**	**11**	**12**
NO	100	98	97	96	93	**0 days/mo**	86	84	79	76	70
YES	0	2	3	4	7	**1 to 2**	5	5	4	6	6
						3 to 5	2	2	2	2	3
						6 to 10	1	2	2	2	2
						11 to 19	2	1	2	2	3
						20 to 29	2	2	4	4	4
						Daily	2	4	7	8	12

Total observations: 18,000.

Table 7.4 reports year-to-year transitions in the smoking outcome for the last five years of the survey. The first five years of smoking transitions are not included because the format of the question changed after year five.[2] Table 7.4 presents a few interesting observations:

- Row 1: 90% of non-smokers will not smoke in the next period.
- Row 7: 81% of daily and 76% of 20-29 days/mo smokers will continue to smoke at least that much in the next period.
- Rows 3 and 4: There is a large dispersion in smoking behavior for individuals who will smoke between 3 and 10 days per month. If the two categories are combined, roughly 20% of these individuals smoke the same amount in the next period while the remaining subsample splits almost evenly between for smoking more and less frequently.

[2]It would be difficult to translate the response "Yes, I have smoked in the last year" in year five into a days-per-month intensity as in year six. There is no discernible difference in year six smoking tendencies between year five smokers and year five non-smokers, primarily due to the small number of year five self-reported smokers (140).

Table 7.4: Smoking Transitions, Conditional on Lagged Use, Percent of Sample (Grades 8-12)

		Period $t+1$ Smoking Habits (days/mo)							$t+1$ relative to t			
		0	**1-2**	**3-5**	**6-10**	**11-19**	**20-29**	**Daily**	**Less**	**Same**	**More**	N
Condi-	**0**	90	4	2	1	1	1	1	–	90	10	4,590
tional on	**1-2**	48	16	8	7	5	6	10	48	16	36	270
Period t	**3-5**	24	16	19	6	9	12	14	40	19	41	129
Smoking	**6-10**	20	13	5	14	13	20	15	38	14	48	79
Habits,	**11-19**	13	4	5	4	22	30	22	26	22	52	97
days per	**20-29**	6	4	4	4	6	34	42	24	34	42	158
month	**Daily**	3	1	0	2	3	10	81	19	81	–	283

Each row is conditional on lagged smoking habits. Total observations: 5,606.

7.4 Attitudinal Variable Definitions

Participants responded to hundreds of subjective questions each year. Most of these questions offer a statement and ask the child to respond using either a yes/no answer, a 5-point Likert scale response, or a word indicating how frequently the statement is true. For the questions considered in the data application, the data record the actual response. Ideally, the dataset would contain questions that uncover additional information about the individual that might directly relate to smoking behavior.[3] Unfortunately, the NGHS survey was not designed with this objective in mind. Instead, I select four attitudinal responses that are influenced by unobserved events and personality traits that might also affect the child's smoking decision. Table 7.5 reports the summary statistics. For estimation, alternative 1 and alternative 3 each combine two response categories.

[3]For example, questions such as "Do you believe that smoking cigarettes is dangerous/cool/a way to lose weight?"

Table 7.5: Attitudinal Response Variables Included in the Data Application

Variable	Description
Att1	Over the last 30 days, ... **I was upset because of something I did not expect.**
Att2	Over the last 30 days, ... **I felt that my problems were becoming so big that I could not handle them.**
Att3	Suppose somebody your age was mean to you or threatened you, or something bad happened to you. **I would stand my ground and fight for what I wanted.**
Att4	Suppose somebody your age was mean to you or threatened you, or something bad happened to you. **I would let my feelings out.**

Alternative	Response	Att1	Att2	Att3	Att4
1	Never OR Once in a While	54%	68%	25%	25%
2	Sometimes	32%	18%	28%	28%
3	Often OR Very Often	14%	14%	47%	47%

$\mathrm{Corr}(\mathrm{Att1}, \mathrm{Att2}) = 0.29$. $\mathrm{Corr}(\mathrm{Att3}, \mathrm{Att4}) = 0.19$.

Other correlation coefficient magnitudes are under 0.05.

The summary statistics are aggregated across years in which attitudinal data is available.

Total Observations: 8,671.

7.5 Other Characteristics

Tables 7.6 and 7.7 and Figures 7.1 and 7.2 report child, parent, and household summary statistics from the initial sample in 1987. Responses in subsequent years are largely unavailable because these questions were found only in the parent surveys, which did not have high response rates after the first year. If more than one parent completed the survey in a particular year, I use the maximum reported value for variables describing the parent's highest education achieved, alcohol use, and smoking habits. In my application, it is more important to know, for example, that a current smoker lives in the household than to know precisely which parent currently smokes. Thus, the percentages for these three variables are biased towards higher categories. For comparison, the national averages from 1987 are given where available. Figure 7.3 reports the price of a pack of cigarettes for each city over time.

Figure 7.1:
Household Composition: Birth Order

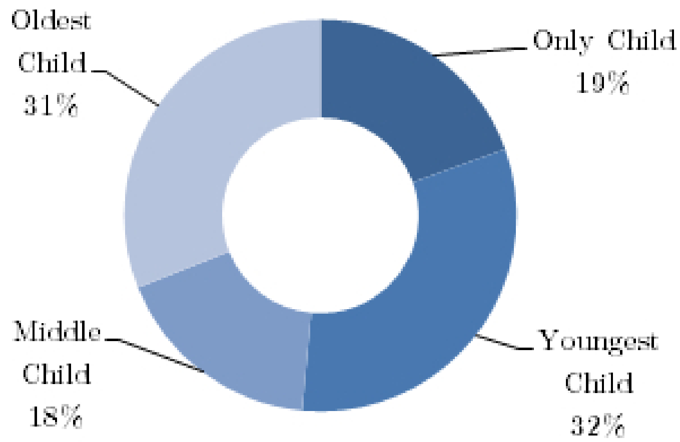


Figure 7.2:
Household Composition: Number of Siblings

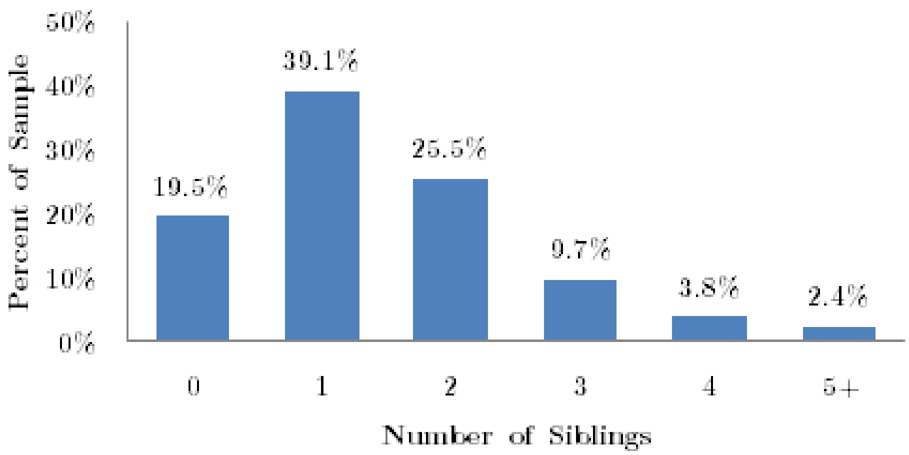


Figure 7.3: Cigarette Prices by Grade Year

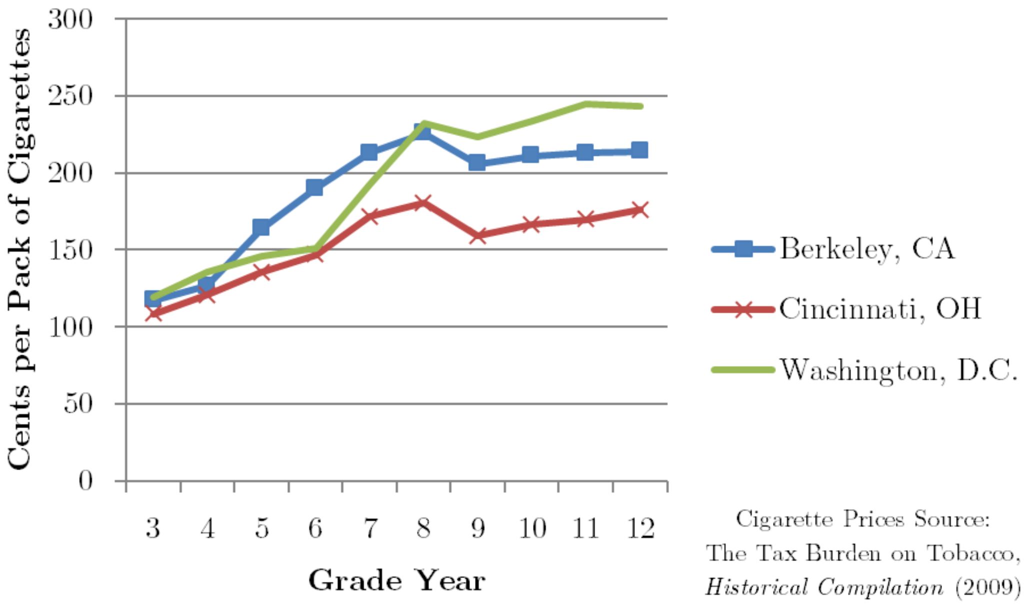


Table 7.6: Child's Characteristics: Initial Sample

Variable	Category	Percent	Description
Race	Black	51%	
	White	49%	
Age			Mean = 10.03. Standard deviation = 0.56.
Health Problem	Yes	14%	Child health problems are reported by the parent. These include asthma, diabetes, high blood pressure, high cholesterol, heart problems, and thyroid problems.
Height	Short for Her Age	13%	The child is "Short for Her Age" if her height is less than 3 inches below the NHANES III average (adjusted for age) and "Tall for Her Age" if her height is greater than 3 inches above the NHANES III average.
	Average Height	69%	
	Tall for Her Age	18%	

Table 7.7: Parent and Household Characteristics: Initial Sample

Variable	Category	Percent	U.S. Avg.
	Berkeley, CA	37%	
Geographical Location	Cincinnati, OH	27%	
	Washington, D.C.	36%	
Number of Parents in Household	1	31%	27%
	2	69%	73%
	HS or less	26%	55%
Parent Education	Some College	39%	19%
	College Grad +	35%	20%
	Under 10,000	42%	
Average Household Income Among	10k - 20k	26%	$18,100
Single-Parent Households	20k - 40k	25%	
	Over 40,000	7%	
	Under 10,000	8%	
Average Household Income Among	10k - 20k	9%	
Two-Parent Households	20k - 40k	33%	$35,600
	Over 40,000	50%	
Lives with Biological Mother	Yes	94%	
Lives with Biological Father	Yes	87%	
	< 1	67%	
Parent: Alcohol, Days per Week	1 - 3	22%	
	4 +	11%	
Parent: Current Smoker	Yes	37%	29%

Variable	Mean	Std. Dev.	% Missing
Parent's Age − Child's Age	27.27	5.71	36%
Parent's BMI	27.29	6.57	27%

This table represents the initial sample; only data from year one (1987) is included.

The 1987 Bureau of Labor Statistics Survey provides the U.S. household income averages.

The 1990 U.S. Census provides the U.S. education averages.

The 1987 Surgeon General Annual Report provides the U.S. smoking averages.

Figure 7.4 displays weight status information for the estimation sample. I use the BMI-for-age statistics from the National Health and Nutritional Examination Survey (NHANES III, 1988-1991) to classify a child as underweight (under the 10th percentile of BMI-for-age) or overweight (over the 85th percentile of BMI-for-age). By comparison, the NGHS sample includes more overweight individuals; each year, roughly 30% of the sample reported a BMI-for-age higher than the 85th percentile of the NHANES sample. A significant portion from each weight status group changes categories from one year to the next, as shown by the weight dynamics statistics in Table 7.8.

Figure 7.4: Weight Status, by Grade Year

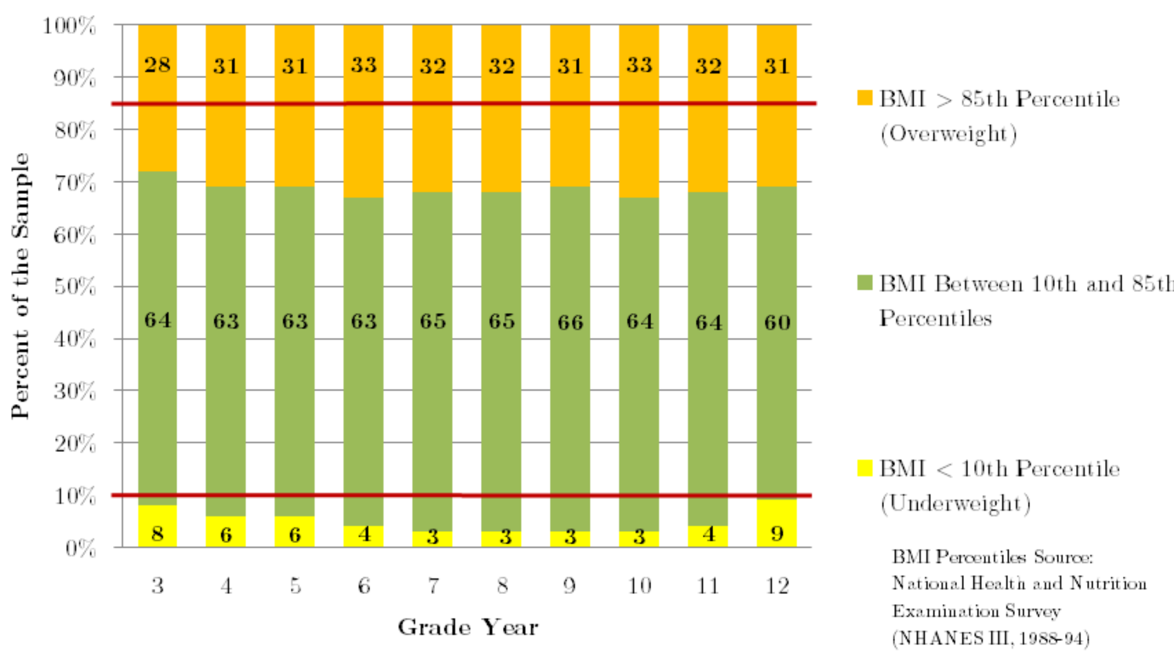


Table 7.8: Weight Status Dynamics, Percent of Sample

		Period $t+1$ Weight Status			
		Underweight	Healthy Weight	Overweight	***N***
Conditional on	Underweight	62	37	1	806
Period t	Healthy Weight	3	89	8	10,781
Weight Status	Overweight	1	15	84	5,125

This table reads from left to right. Total observations: 16,712.

7.6 Accuracy of Data

With all self-reported data, the researcher must rely on participants to provide honest and accurate responses. There might be some concern over the legitimacy of the NGHS data because the study asked teenagers to answer subjective questions. For example, the girl might select the option that she *wants* to be true, perhaps out of embarrassment or guilt, but the NGHS takes measures to prevent this behavior. Upon arrival to the facility, the child is separated from her parents to complete the physical examination and written survey. On several survey forms, the child is told that her answers are confidential and will not be shared with parents, teachers, or friends. In particular, this reminder appears directly above questions related to smoking habits.

In the sixth year of the study, a blood analysis reported the girls' cotinine levels. According to the Foundation for Blood Research, cotinine can only be produced from nicotine, which means the child's cotinine level is an indicator that the individual has smoked recently. Cotinine is measured in nanograms per milliliter (ng/mL) of blood and can be detected in the blood within hours of smoking. However, if an individual does not smoke for ten days, her cotinine level will return to that of a non-smoker, resulting in a false-negative. Additionally, analysis of a non-smoker's blood sample might report non-zero levels of cotinine (false-positive) if she is exposed to secondhand smoke or suffers from heart disease. Though this measure is not perfect, the consistency between self-reported smoking and measured cotinine level gives some idea of the data's accuracy. Scientific studies typically find that blood samples from frequent smokers contain at least 10 ng of cotinine per mL. Thus, we would expect individuals who never smoke to report levels less than 10 ng/mL and individuals who smoke regularly to report levels greater than 10 ng/mL. We cannot form an appropriate hypothesis for infrequent users since cotinine leaves the system in just a few days. Table 7.9 matches the child's self-reported cigarette use with the cotinine level from her blood test. 98% of self-reported non-smokers contain low levels of cotinine in their blood, and 92% of self-reported daily smokers contain high levels of cotinine in their blood. The results are also consistent for individuals who smoke one to two days per month (92% low levels) and 20 to 29 days per month (77% high levels).

Table 7.9: Cotinine Level, Conditional on Smoking Habits, Percent of Sample (Grade 9)

		Conditional on Smoking Habits (days/mo)						
		0	**1-2**	**3-5**	**6-10**	**11-19**	**20-29**	**Daily**
Cotinine (ng/mL)	**0**	95	86	66	63	26	14	8
	1 − 9	3	6	14	16	13	9	0
	10 − 99	1	6	3	16	39	37	17
	≥ 100	1	2	17	5	22	40	75
	N	1,174	70	35	19	23	35	64

Samples from individuals who smoke frequently typically contain at least 10 ng of cotinine per mL of blood. The percentages in each cell are conditional on reported smoking habits. This table reads from top to bottom. Total observations: 1,420.

Response accuracy can be measured for questions that were completed by both the child and the clinician, who has no incentive to misreport. In response to the statement "Right now I look like...," the child selected among nine body images, shown in Figure 7.5. In the first five years of the study, the NGHS clinician also completed this question. Because there were only subtle differences between adjacent body images, a minor discrepancy between responses might be due to interpretation of the images instead of intentional misreporting. Table 7.10 shows that 40% of the responses match exactly, and the difference between 86% of the response pairs is at most one category. While this finding does not indicate that responses to every question on the survey are this accurate, it suggests that the girls' responses to this subjective question are fairly accurate.

Figure 7.5:
"Right now I look like ______"

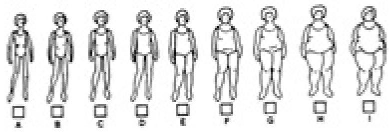


Table 7.10: Discrepancy in Responses

Absolute Difference:	0	1	2	3+
Percent of Sample:	40	46	12	2

Total observations: 10,877.

Chapter 8

Results from the Data Application

I separately estimate **Model I**, **Model E**, and **Model C** using the NGHS estimation sample described in Chapter 7. The outcome of interest, smoking behavior, is transformed into a binary variable that equals one if the individual reports any positive smoking frequency. The four attitudinal response variables summarized in Table 7.5 are omitted from **Model I**, included as explanatory variables in **Model E**, and incorporated as outcomes in **Model C**.

8.1 Unobserved Heterogeneity

Two latent-factor distributions are approximated by the empirical model; one approximates permanent unobserved heterogeneity and one approximates time-varying unobserved heterogeneity.[1] **Model I** successfully identifies three mass points for the permanent latent factor distribution and three mass points for the time-varying latent factor distribution, **Model E** successfully identifies two permanent and two time-varying mass points, and **Model C** successfully identifies four permanent and four time-varying mass points. Tables 8.1, 8.2, and 8.3 report the estimated points of support and respective and probabilities for each mass point of the latent factors included in the models.[2]

---

[1]To determine the optimal number of mass points for each model, I choose the specification that estimates the most points of support without returning a zero-probability mass point.

[2]For tables that display regression results, an asterisk denotes that an estimated coefficient is statistically significant at the 0.05 significance level.

Table 8.1: **Model I** Estimated Latent-Factor Distribution

		Permanent			Time-Varying		
		λ_1	λ_2	λ_3	ν_{1t}	ν_{2t}	ν_{3t}
	Probability (%)	3	45	52	2	6	92
Smoking	Point Estimate	0	-6.36	-3.21	0	-7.56	-1.06
	Standard Error	–	*0.50*	*0.44*	–	*2.85*	*0.78*
	P-Value	–	0.00*	0.00*	–	0.01*	0.17
Attrition	Point Estimate	0	-1.77	-0.52	0.00	-55.71	-29.81
	Standard Error	–	*0.73*	*0.45*	–	*358.16*	*154.57*
	P-Value	–	0.02*	0.25	–	0.88	0.85

Mass points λ_1 and ν_{1t} are normalized to zero.

Relative to alternative "No Smoking" or "Does Not Attrit."

Table 8.2: **Model E** Estimated Latent-Factor Distribution

		Permanent		Time-Varying	
		λ_1	λ_2	ν_{1t}	ν_{2t}
	Probability (%)	57	43	36	64
Smoking	Point Estimate	0	-2.55	0	-8.93
	Standard Error	–	*0.28*	–	*6.32*
	P-Value	–	0.00*	–	0.16
Attrition	Point Estimate	0	-1.00	0.00	-0.67
	Standard Error	–	*0.44*	–	*2.18*
	P-Value	–	0.02*	–	0.76

Mass points λ_1 and ν_{1t} are normalized to zero.

Relative to alternative "No Smoking" or "Does Not Attrit."

Table 8.3: **Model C** Estimated Latent-Factor Distribution

		Permanent Mass Points				Time-Varying Mass Points			
		λ_1	λ_2	λ_3	λ_4	ν_{1t}	ν_{2t}	ν_{3t}	ν_{4t}
	Probability (%)	26	27	20	22	19	58	10	13
Smoking	Point Estimate	0	-0.25	-0.43	-0.43	0	-0.22	0.72	0.72
	Standard Error	–	*0.18*	*0.16*	*0.17*	–	*0.28*	*0.33*	*0.30*
	P-Value	–	0.16	0.01*	0.01*	–	0.43	0.03*	0.02*
Att1 = 2	Point Estimate	0	-1.07	-0.84	-0.81	0	-1.92	-0.92	0.43
	Standard Error	–	*0.16*	*0.13*	*0.14*	–	*0.26*	*0.35*	*1.04*
	P-Value	–	0.00*	0.00*	0.00*	–	0.00*	0.01*	0.68
Att1 = 3	Point Estimate	0	-2.02	-1.78	-1.97	0	-1.83	0.24	3.02
	Standard Error	–	*0.30*	*0.22*	*0.24*	–	*0.47*	*0.53*	*1.15*
	P-Value	–	0.00*	0.00*	0.00*	–	0.00*	0.65	0.01*
Att2 = 2	Point Estimate	0	-1.79	-2.07	-1.25	0	-1.40	-0.77	0.93
	Standard Error	–	*0.21*	*0.18*	*0.23*	–	*0.22*	*0.27*	*0.40*
	P-Value	–	0.00*	0.00*	0.00*	–	0.00*	0.00*	0.02*
Att2 = 3	Point Estimate	0	-2.68	-3.10	-2.08	0	-0.76	0.40	3.11
	Standard Error	–	*0.33*	*0.27*	*0.36*	–	*0.42*	*0.42*	*0.47*
	P-Value	–	0.00*	0.00*	0.00*	–	0.07	0.33	0.00*
Att3 = 2	Point Estimate	0	0.19	0.03	-0.66	0	-1.40	-0.76	-1.47
	Standard Error	–	*0.18*	*0.16*	*0.15*	–	*0.25*	*0.33*	*0.30*
	P-Value	–	0.31	0.85	0.00*	–	0.00*	0.02*	0.00*
Att3 = 3	Point Estimate	0	0.61	0.44	-2.33	0	-1.04	0.54	-1.00
	Standard Error	–	*0.27*	*0.21*	*0.25*	–	*0.28*	*0.43*	*0.34*
	P-Value	–	0.02*	0.03*	0.00*	–	0.00*	0.21	0.00*

Mass points λ_1 and ν_{1t} are normalized to zero.

Relative to alternative “No Smoking,” “Att* = 1,” or “Does Not Attrit.”

Continued on the next page....

Table 8.3 Continued: **Model C** Estimated Latent-Factor Distribution

		Permanent Mass Points				Time-Varying Mass Points			
		λ_1	λ_2	λ_3	λ_4	ν_{1t}	ν_{2t}	ν_{3t}	ν_{4t}
Att4 = 2	Point Estimate	0	-0.33	0.55	-0.09	0	-1.33	-2.70	-1.69
	Standard Error	–	*0.16*	*0.18*	*0.14*	–	*0.25*	*2.88*	*0.29*
	P-Value	–	0.04*	0.00*	0.52	–	0.00*	0.35	0.00*
Att4 = 3	Point Estimate	0	-1.93	1.18	-1.56	0	-0.43	6.27	-0.80
	Standard Error	–	*0.43*	*0.25*	*0.25*	–	*0.35*	*6.11*	*0.47*
	P-Value	–	0.00*	0.00*	0.00*	–	0.22	0.31	0.09
Attrition	Point Estimate	0	0.37	-0.16	0.00	0	-0.28	0.12	-0.08
	Standard Error	–	*0.18*	*0.18*	*0.20*	–	*0.25*	*0.31*	*0.35*
	P-Value	–	0.04*	0.38	0.99	–	0.28	0.70	0.81

Mass points λ_1 and ν_{1t} are normalized to zero.

Relative to alternative "No Smoking," "Att* = 1," or "Does Not Attrit."

The theoretical model of behavior shows that several unobserved factors simultaneously influence the outcome of interest (smoking) and the attitudinal responses, suggesting that the error terms across outcomes are correlated. The results for **Model C** in Table 8.3 confirm this claim, as evidenced by the statistically significant latent-factor point estimates in each equation. For example, observe the relationship between the dependent variables Smoking and Att1 ("upset"). Mass points λ_3 and λ_4 show a positive correlation between the error terms because the point estimates share the same sign.[3] These mass points suggest that some unobserved factors simultaneously discourage smoking and reporting higher frequencies of feeling upset (Att1 $=$ 2 or 3). Mass point ν_{3t}, which has a positive point estimate in the Smoking equation $\left(\hat{\nu}_{3t}^{\text{Smoke}} = 0.72\right)$ and negative point estimates in the Att1 equation $\left(\hat{\nu}_{3t}^{\text{Att1=2}} = -0.92 \text{ and } \hat{\nu}_{3t}^{\text{Att1=3}} \text{ is statistically insignificant from zero}\right)$, suggests a negative correlation between these variables' unobserved factors. However, generally speaking, there is a positive correlation among the unobserved factors that influence the Smoking and Att1 outcomes; only mass point ν_{3t}, which has a low predicted probability (10%), suggests that a negative correlation exists.

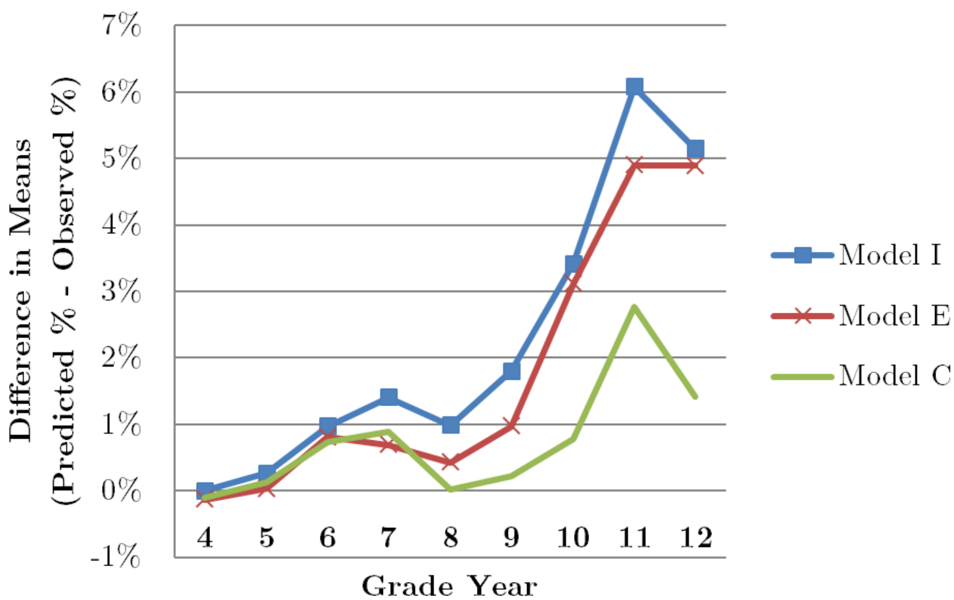
The model would reveal to the researcher if the outcome of interest and the attitudinal response are indeed conditionally independent. Observe the estimated distribution for time-varying mass point ν_{2t}. The coefficient in the smoking equation is insignificantly different from zero. However, most of the coefficients for ν_{2t} in the attitudinal response equations share the same sign and are significantly different from zero. Thus, mass point ν_{2t} suggests that there exist unobserved factors that are common across attitudinal outcomes but uncorrelated with the error term of the smoking equation. The signs on these parameters suggest a positive correlation among the error terms in the attitudinal equations. If this were true for *every* mass point, the model would suggest that the outcome of interest and attitudinal response were conditionally independent. However, the estimates in Table 8.3 for mass points λ_3, λ_4, ν_{3t}, and ν_{4t} show that this is not the case.

[3] Table 8.3 shows that the point estimates are all negative: $\left\{\hat{\lambda}_3^{\text{Smoke}}, \hat{\lambda}_3^{\text{Att1=2}}, \hat{\lambda}_3^{\text{Att1=3}}\right\} = \{-0.43, -0.84, -1.78\}$ and $\left\{\hat{\lambda}_4^{\text{Smoke}}, \hat{\lambda}_4^{\text{Att1=2}}, \hat{\lambda}_4^{\text{Att1=3}}\right\} = \{-0.43, -0.81, -1.97\}$.

8.2 Model Predictions

Without knowing the true parameters of the model, bias and efficiency comparisons across models can be misguided. Instead, for the data application, the accuracy of the models' predictions are compared to test each model's performance. Figure 8.1 and Table 8.4 report the difference between the sample average of the observed smoking outcomes and the sample average of the predicted smoking probabilities for each model by grade year.[4,5] The smoking predictions from **Model C** most accurately match the observed smoking outcomes reported in the data, especially for later years during which a greater percentage of the sample chooses to smoke. In addition, the standard errors of the predicted means are smaller for **Model C**. This evidence suggests that **Model C** outperforms **Model I** and **Model E**.

Figure 8.1: Difference Between Observed and Predicted Smoking Outcome Means, by Grade Year



[4]The full prediction results are available in Tables A.1, A.2, and A.3.

[5]Only individuals who report smoking behavior in the data sample are used for the simulations. Smoking behavior is simulated forward, starting in grade 4, and the smoking history variables for subsequent grade years comprise the simulated outcomes from previous years. In grade 3, the initial condition for this dynamic model, no one reports that they smoke. Thus, there is no need to jointly estimate an initial conditions equation because there is no variance in the dependent variable for grade 3.

Figure 8.2: Difference Between Observed and Predicted, with Confidence Intervals

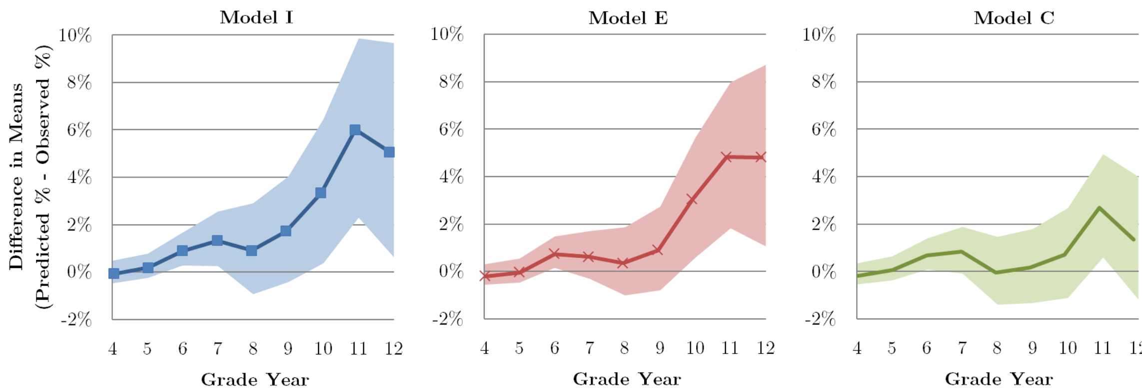


Table 8.4: Difference Between Observed and Predicted Smoking Outcome Means, by Year

	Grade Year →	4	5	6	7	8	9	10	11	12
True	Pr(Smoke)	1.56	2.55	3.66	7.32	13.59	16.54	21.10	24.31	30.05
Model I	$\widehat{\Pr}$(Smoke)	1.56	2.81	4.64	8.72	14.58	18.34	24.51	30.39	35.19
	Standard Error	*0.24*	*0.26*	*0.35*	*0.59*	*0.98*	*1.14*	*1.55*	*1.93*	*2.31*
	Difference	**0.00**	**0.26**	**0.98**	**1.40**	**0.99**	**1.80**	**3.41**	**6.08**	**5.14**
Model E	$\widehat{\Pr}$(Smoke)	1.64	2.85	4.87	8.70	15.01	18.46	25.29	30.40	36.24
	Standard Error	*0.94*	*1.26*	*1.56*	*1.98*	*2.42*	*2.79*	*3.16*	*3.45*	*3.72*
	Difference	**0.08**	**0.30**	**1.21**	**1.38**	**1.42**	**1.92**	**4.19**	**6.09**	**6.19**
Model C	$\widehat{\Pr}$(Smoke)	1.45	2.67	4.40	8.21	13.62	16.76	21.88	27.08	31.46
	Standard Error	*0.23*	*0.26*	*0.33*	*0.50*	*0.73*	*0.80*	*0.97*	*1.11*	*1.32*
	Difference	**-0.11**	**0.12**	**0.74**	**0.89**	**0.03**	**0.22**	**0.78**	**2.77**	**1.41**

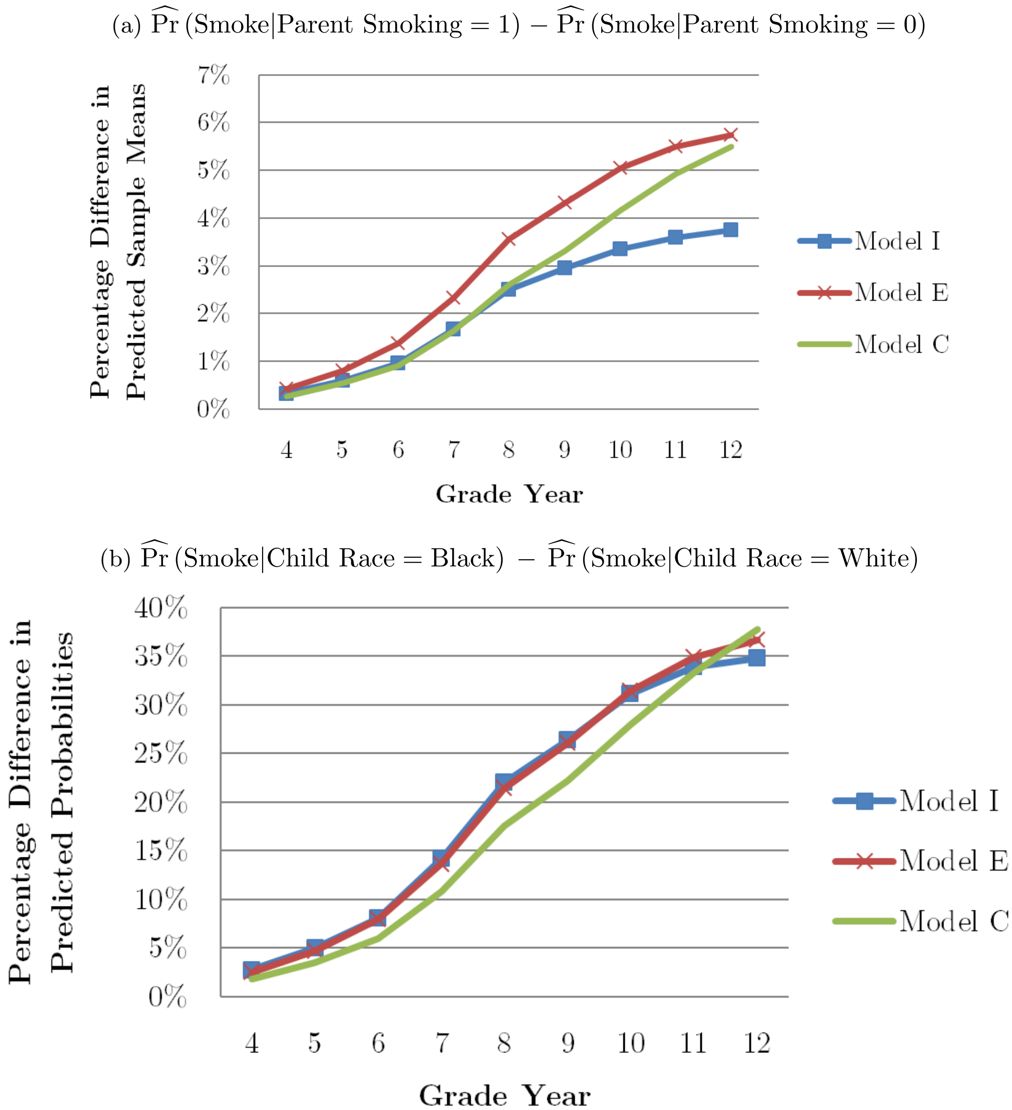
The statistics in this table were generated by simulating smoking behavior 200 times for each model.

8.3 Counterfactual Experiments

I conduct three counterfactual experiments using the estimated parameters to show that these models generate different policy forecasts. To weigh the costs and benefits of, for example, a marketing campaign to discourage smoking, the marginal effects of key variables must be measured accurately. First, I estimate the marginal effect of the parents' smoking habits by replacing each parents' response with "0 − Not a Current Smoker," predicting the child's smoking behaviors, and repeating the process with "1 − Current Smoker." I perform the same experiment for race to calculate the difference in predicted probabilities between white and black girls. The difference between sample average predicted probabilities is summarized by Figure 8.3. These two experiments suggest that anti-smoking marketing campaigns would be more effective if they target white girls who live with a parent that smokes. Figure 8.3a suggests that, compared to **Model C**, **Model I** understates the marginal effect of a smoker living in the child's household and **Model E** overstates this effect. In Figure 8.3b, **Models I** and **E** predict a higher probability of smoking for white girls in the first 8 years of the sample compared to **Model C**. For some grades, the difference is greater than 4%.

Figure 8.3: Estimated Marginal Effects:

Parent Smoking Habits and Child Race, Sample Means, by Grade

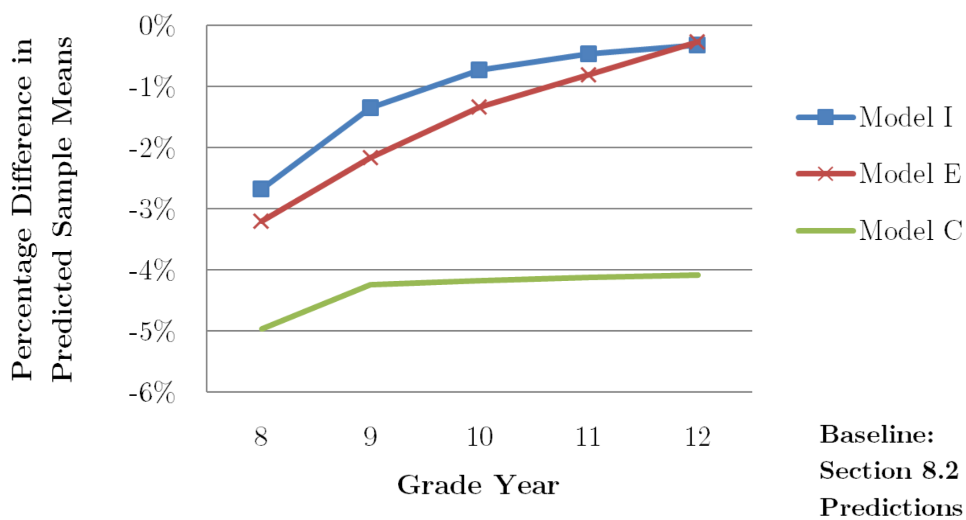


Finally, I estimate the long-term impact of choosing not to smoke before the eighth grade. For grades 4 through 7, I manually overwrite each individual's responses with "0 − No Smoking." The model simulates smoking behavior for grade years 8 to 12. Figure 8.4 reports the difference between the predicted outcomes from Section 8.2, which uses the actual data, and the predicted outcomes from this experiment. These percentages answer the question "What would have happened to the sample average smoking behavior if all NGHS participants had abstained from smoking before the eighth grade?" There are two noticeable differences between

Model C and **Models I** and **E**. First, **Models I** and **E** understate the estimated impact of this abstinence compared to **Model C**. Second, the sample averages follow different trends. **Models I** and **E** predict that a significant portion of the sample will abstain from smoking initially, but the long-term effectiveness fades over time. For **Model C**, the predicted sample mean is roughly 4% lower in *each* of the remaining survey periods. The marginal effect of abstaining before eighth grade is a function of the estimated coefficients for variables describing the individual's smoking history. Table A.4 in the **Appendix** shows that the estimated response parameters from **Model C** differ substantially from the estimated response parameters from **Models I** and **E**, which explains the significant difference between the simulated marginal effects depicted in Figure 8.4.

Figure 8.4: Difference Between Predicted Sample Average Smoking Behavior

$$\widehat{\Pr}\,(\text{Smoke}|\text{Abstain from smoking before 8th grade}) \; - \; \widehat{\Pr}\,(\text{Smoke})$$



The addition of attitudinal responses as outcome variables in **Model C** enhances the model's ability to approximate the distribution of unobserved heterogeneity. From Figure 8.2, I find that **Model C** generates more accurate and more precise predictions of consumer behavior compared to specifications that either ignore the attitudinal data (**Model I**) or incorporate attitudinal responses as explanatory variables (**Model E**). I conduct three counter-factual experiments to estimate the marginal effects of three explanatory variables: parent's smoking habits, child's weight status, and child's smoking history. Because Figure 8.2 suggests that

Model C outperforms **Models I** and **E**, I show in Figures 8.3 and 8.4 that **Models I** and **E** will either over- or under-state the true marginal effects of the explanatory variables on the consumer choice probabilities.

Chapter 9

Conclusion

This research supports the use of attitudinal information as outcome variables in empirical work by investigating the relationship between an individual's decisions and attitudinal responses. The theoretical model of behavior suggests that a model which jointly estimates the decisions and attitudinal responses as outcomes most appropriately captures the relationship between these simultaneously-determined outcomes. The econometric proof shows that this model increases the efficiency of the estimated parameters compared to a model that ignores the attitudinal responses. The Monte Carlo experiment provides evidence that this model reduces both the bias and standard errors of the estimated parameters compared to models that either ignore or inappropriately incorporate attitudinal responses. The data application using the NGHS dataset demonstrates that more accurate predictions of behavior are produced with this model, an improvement that is valuable in counter-factual forecasting experiments.

For further research with the NGHS dataset, I will incorporate more data described in Chapter 7, such as the individual's food consumption, exercise habits, and drinking behavior, into the model as additional jointly-estimated outcomes. Also, the data application can be enhanced by including more than just four attitudinal response outcomes. Other datasets that already include attitudinal information are excellent candidates for the use of the empirical specification in this paper. For research studies that will collect their own data, adding just a few suitable attitudinal questions to the survey could be a low-cost way to enhance the model's accuracy, efficiency, and predictive power.

Appendix A

Appendix

A.1 Chapter 3.2: Primitives in the Optimal Demand Functions

The solution to the model shows that the optimal decision is a function of primitive parameters $\boldsymbol{\mu}$, $\boldsymbol{\beta}$, and $\boldsymbol{\alpha}$. For notational ease, I impose the assumption that the utility function is additively separable in consumption of y_{it} and c_{it}. Further, I assume that individuals know how the history variable Y_{it} evolves; that is, $Y_{it} = Y\left(y_{it-1}, Y_{it-1}\right)$. The objective function becomes:

$$\begin{aligned} \Psi^S\left(\boldsymbol{\mu}, \boldsymbol{\beta}, \boldsymbol{\alpha}, X_{it}, Y_{it}, \epsilon_{it}^S\right) = \quad & \max_y \quad U^S(y, [I_{it} - p_t y]; \boldsymbol{\mu}, X_{it}, Y_{it}, \epsilon_{it}^{yS}) \\ & + \boldsymbol{\beta} * \pi_1(\boldsymbol{\alpha}, \bullet) * V^1\left(\boldsymbol{\mu}, \boldsymbol{\beta}, \boldsymbol{\alpha}, X_{it+1}, Y\left(y, Y_{it}\right)\right) \\ & + \boldsymbol{\beta} * \pi_0(\boldsymbol{\alpha}, \bullet) * V^0\left(\boldsymbol{\mu}, \boldsymbol{\beta}, \boldsymbol{\alpha}, X_{it+1}, Y\left(y, Y_{it}\right)\right). \end{aligned} \tag{A.1}$$

The first order condition with respect to the individual's decision y_{it} is:

$$\begin{aligned} 0 = \tfrac{\partial U^S(\bullet_t)}{\partial y} - \tfrac{\partial U^S(\bullet_t)}{\partial c} p_t \quad & + \boldsymbol{\beta}\left(\tfrac{\partial \pi_1(\boldsymbol{\alpha}, \bullet)}{\partial y} V^1\left(\bullet_{t+1}\right) + \pi_1\left(\boldsymbol{\alpha}, \bullet\right) \tfrac{\partial V^1(\bullet_{t+1})}{\partial Y} * \tfrac{\partial Y}{\partial y}\right) \\ & + \boldsymbol{\beta}\left(\tfrac{\partial \pi_0(\boldsymbol{\alpha}, \bullet)}{\partial y} V^0\left(\bullet_{t+1}\right) + \pi_0\left(\boldsymbol{\alpha}, \bullet\right) \tfrac{\partial V^0(\bullet_{t+1})}{\partial Y} * \tfrac{\partial Y}{\partial y}\right) \end{aligned} \tag{A.2}$$

Solving for y_{it} results in the individual's optimal decision y_{it}^*, which is a function of the primitive parameters:

- The expected transition probabilities $\boldsymbol{\pi}$ are functions of the subjective-expectations operator $\boldsymbol{\alpha}$. Multiplied by these probabilities are two non-zero elements: the partial derivative of the maximal expected value of each state with respect to the individual's consumption history $(\partial V / \partial Y)$ and the derivative of the evolution equation $(\partial Y / \partial y)$.[1] Therefore,

[1] The individual's history of consumption will impact future utility streams; thus, $\left(\frac{\partial V(\bullet_{t+1})}{\partial Y}\right) \neq 0$. Also, because Y_{it+1} is a function of y_{it}, $\frac{\partial Y}{\partial y} \neq 0$.

parameter $\boldsymbol{\alpha}$ will remain in the solution for y_{it}^{*}.[2]

- In the theoretical model, I assume that the marginal utility of consuming good y, $\left(\frac{\partial U^S(\bullet_t)}{\partial y}\right)$, is a function of parameter $\boldsymbol{\mu}$, which means that parameter $\boldsymbol{\mu}$ will remain in the solution for y_{it}^{*}.

- The terms in parentheses do not equal zero. Therefore, y_{it}^{*} is a function of discount factor $\boldsymbol{\beta}$.

The theoretical model shows that these primitive parameters influence the optimal demand for y; that is,

$$y_{it}^{*} = y^{*}\left(\boldsymbol{\mu}, \boldsymbol{\beta}, \boldsymbol{\alpha},\ \bullet\right). \tag{A.3}$$

A.2 Chapter 4: Empirical Model with State Transitions for S_{it}

The empirical model incorporates state variable S_{it} by estimating the transition probabilities for S_{it} and allowing S_{it} to impact the choice probabilities. To construct a basic model, I consider a binary variable S_{it}, normalize the parameters with respect to outcome $S_{it} = 0$, and incorporate latent factors.[3] An assumption that the idiosyncratic error term for the transition probabilities are *iid* from a *GEV* distribution allows for the following closed form expression:

$$\Pr\left(S_{it+1} = 1 | y_{it}, X_{it}, Y_{it}, S_{it}, \lambda^S, \nu_t^S\right) = \frac{\exp\left(y_{it}\rho_y + X_{it}\rho_X + Y_{it}\rho_Y + S_{it}\rho_S + \lambda^S + \nu_t^S\right)}{1 + \exp\left(y_{it}\rho_y + X_{it}\rho_X + Y_{it}\rho_Y + S_{it}\rho_S + \lambda^S + \nu_t^S\right)} \tag{A.4}$$

In the theoretical model, the marginal effects of X_{it} and Y_{it} vary across realizations of S_{it}. The empirical model incorporates this feature by interacting S_{it} with the covariates in the Taylor Series expansions of outcome probabilities for y_{it} and r_{it}.

$$\Pr\left(y_{it} = y | X_{it}, Y_{it}, S_{it}, \lambda^y, \nu_t^y\right) = \frac{\exp\left(X_{it}\gamma_X^y + X_{it}S_{it}\gamma_{XS}^y + Y_{it}\gamma_Y^y + Y_{it}S_{it}\gamma_{YS}^y + \lambda^y + \nu_t^y\right)}{\sum_{\tilde{y}} \exp\left(X_{it}\gamma_X^{\tilde{y}} + X_{it}S_{it}\gamma_{XS}^{\tilde{y}} + Y_{it}\gamma_Y^{\tilde{y}} + Y_{it}S_{it}\gamma_{YS}^{\tilde{y}} + \lambda^{\tilde{y}} + \nu_t^{\tilde{y}}\right)} \tag{A.5}$$

$$\Pr\left(r_{it} = r | X_{it}, Y_{it}, S_{it}, \lambda^r, \nu_t^r\right) = \frac{\exp\left(X_{it}\phi_X^r + X_{it}S_{it}\phi_{XS}^r + Y_{it}\phi_Y^r + Y_{it}S_{it}\phi_{YS}^r + \lambda^r + \nu_t^r\right)}{\sum_{\tilde{r}} \exp\left(X_{it}\phi_X^{\tilde{r}} + X_{it}S_{it}\phi_{XS}^{\tilde{r}} + Y_{it}\phi_Y^{\tilde{r}} + Y_{it}S_{it}\phi_{YS}^{\tilde{r}} + \lambda^{\tilde{r}} + \nu_t^{\tilde{r}}\right)} \tag{A.6}$$

---

[2] To ensure that the $\boldsymbol{\alpha}$ remains, I assume that $V^1\left(\bullet_{t+1}\right) \neq V^0\left(\bullet_{t+1}\right)$ and $\frac{\partial V^1(\bullet_{t+1})}{\partial Y} \neq \frac{\partial V^0(\bullet_{t+1})}{\partial Y}$. If this were not the case, the state variable would be irrelevant. In addition, I rule out the highly unlikely case that $\frac{\partial \pi_1(\boldsymbol{\alpha},\bullet)}{\partial y} V^1\left(\bullet_{t+1}\right) + \pi_1\left(\boldsymbol{\alpha}, \bullet\right) \frac{\partial V^1(\bullet_{t+1})}{\partial Y}\frac{\partial Y}{\partial y} = \frac{\partial \pi_1(\boldsymbol{\alpha},\bullet)}{\partial y} V^0\left(\bullet_{t+1}\right) + \pi_1\left(\boldsymbol{\alpha}, \bullet\right) \frac{\partial V^0(\bullet_{t+1})}{\partial Y}\frac{\partial Y}{\partial y}$.

[3] Expansions might incorporate broader definitions for S_{it} (i.e. a categorical or continuous variable) or heterogeneous effects of y_{it} on S_{it+1} (i.e. a learning process or random parameter for ρ_y within (A.4)).

The likelihood function consists of all three probabilities.

$$\begin{aligned}
\mathfrak{L}\left(\boldsymbol{y}, \boldsymbol{\mathcal{R}}, \boldsymbol{S} | \boldsymbol{Z}; \Theta, \mathfrak{p}, \mathfrak{q}\right) = \quad & \prod_{i=1}^{N} \sum_{k=1}^{K} \mathfrak{p}_k \prod_{t=1}^{T} \sum_{l=1}^{L} \mathfrak{q}_l \\
& * \prod_{y=0}^{\mathrm{Y}} \Pr\left(y_{it} = y | \boldsymbol{z}_{it}, \lambda_k, \nu_{tl}\right)^{\mathbb{1}(y_{it}=y)} \\
& * \prod_{q=1}^{Q} \prod_{r_q=0}^{\mathrm{R}_q} \Pr\left(r_{qit} = r_q | \boldsymbol{z}_{it}, \lambda_k, \nu_{tl}\right)^{\mathbb{1}(r_{qit}=r)} \\
& * \Pr\left(S_{it+1} = 1 | y_{it}, S_{it}, \boldsymbol{z}_{it}, \lambda_k, \nu_{tl}\right)^{\mathbb{1}(S_{it+1}=1)} \\
& * \left(1 - \Pr\left(S_{it+1} = 1 | y_{it}, S_{it}, \boldsymbol{z}_{it}, \lambda_k, \nu_{tl}\right)\right)^{\mathbb{1}(S_{it+1}=0)}
\end{aligned} \tag{A.7}$$

A.3 Chapter 8: Coefficient Estimates and Prediction Results

Table A.1:
Model I: Observed Outcomes and Predicted Probabilities, Sample Means, by Year

Grade Year →	4	5	6	7	8	9	10	11	12
$\widehat{\Pr}$(Smoke)	1.56	2.81	4.64	8.72	14.58	18.34	24.51	30.39	35.19
Standard Error	*0.24*	*0.26*	*0.35*	*0.59*	*0.98*	*1.14*	*1.55*	*1.93*	*2.31*
Pred. − True	**0.00**	**0.26**	**0.98**	**1.40**	**0.99**	**1.80**	**3.41**	**6.08**	**5.14**
$\widehat{\Pr}$(Attrit)	5.94	8.32	10.89	12.84	12.61	10.78	9.08	7.44	–
Standard Error	*3.60*	*3.75*	*3.81*	*3.86*	*3.90*	*3.90*	*3.93*	*3.93*	
Pred. − True	**2.07**	**3.09**	**4.34**	**-1.01**	**5.81**	**4.60**	**2.14**	**3.49**	

Attitudinal responses are not used in **Model I**. Sample Size: 15,621.

Table A.2:
Model E: Observed Outcomes and Predicted Probabilities, Sample Means, by Year

Grade Year →	4	5	6	7	8	9	10	11	12
$\widehat{\Pr}$(Smoke)	1.64	2.85	4.87	8.70	15.01	18.46	25.29	30.40	36.24
Standard Error	*0.94*	*1.26*	*1.56*	*1.98*	*2.42*	*2.79*	*3.16*	*3.45*	*3.72*
Pred. − True	**0.08**	**0.30**	**1.21**	**1.38**	**1.42**	**1.92**	**4.19**	**6.09**	**6.19**
$\widehat{\Pr}$(Attrit)	3.70	7.37	8.08	11.55	9.64	9.60	6.38	5.41	–
Standard Error	*0.75*	*0.93*	*0.96*	*1.06*	*1.06*	*1.10*	*0.92*	*1.05*	
Pred. − True	**-0.17**	**2.14**	**1.53**	**-2.30**	**2.84**	**3.42**	**-0.56**	**1.46**	

Attitudinal responses are used as explanatory variables in **Model E**. Sample Size: 15,621.

Table A.3:
Model C: Observed Outcomes and Predicted Probabilities, Sample Means, by Year

Grade Year →	4	5	6	7	8	9	10	11	12
$\widehat{\Pr}$(Smoke)	1.45	2.67	4.40	8.21	13.62	16.76	21.88	27.08	31.46
Standard Error	*0.23*	*0.26*	*0.33*	*0.50*	*0.73*	*0.80*	*0.97*	*1.11*	*1.32*
Pred. – True	**-0.11**	**0.12**	**0.74**	**0.89**	**0.03**	**0.22**	**0.78**	**2.77**	**1.41**
$\widehat{\Pr}$(Att1=1)	55.55	–	52.26	–	52.61	–	55.21	–	54.79
Standard Error	*1.37*		*1.34*		*1.17*		*1.46*		*1.69*
Pred. – True	**0.54**		**-1.33**		**1.88**		**-0.86**		**1.22**
$\widehat{\Pr}$(Att1=2)	28.33	–	33.49	–	33.05	–	30.94	–	32.44
Standard Error	*0.96*		*1.08*		*0.90*		*1.22*		*1.48*
Pred. – True	**-0.51**		**0.86**		**-1.84**		**1.10**		**-0.71**
$\widehat{\Pr}$(Att1=3)	16.12	–	14.25	–	14.34	–	13.85	–	12.77
Standard Error	*1.01*		*0.92*		*0.86*		*1.02*		*1.05*
Pred. – True	**-0.03**		**0.47**		**-0.04**		**-0.24**		**-0.51**
$\widehat{\Pr}$(Att2=1)	64.44	–	68.57	–	68.90	–	68.08	–	70.65
Standard Error	*1.06*		*1.12*		*1.05*		*1.27*		*1.30*
Pred. – True	**-0.13**		**-1.04**		**-0.09**		**-0.48**		**0.23**
$\widehat{\Pr}$(Att2=2)	17.98	–	18.26	–	17.52	–	17.70	–	17.82
Standard Error	*0.81*		*0.83*		*0.75*		*1.04*		*1.12*
Pred. – True	**-0.18**		**0.72**		**-0.74**		**0.84**		**-0.19**
$\widehat{\Pr}$(Att2=3)	17.58	–	13.17	–	13.59	–	14.22	–	11.54
Standard Error	*0.89*		*0.82*		*0.76*		*0.96*		*0.94*
Pred. – True	**0.31**		**0.32**		**0.85**		**-0.35**		**-0.03**

Attitudinal responses are jointly estimated as outcomes in **Model C**.

Attitudinal responses are only available every other year. Sample Size: 15,621.

Continued on the next page....

Table A.3 Continued:

Model C: Observed Outcomes and Predicted Probabilities, Sample Means, by Year

Grade Year →	4	5	6	7	8	9	10	11	12
$\widehat{\Pr}$(Att3=1)	46.39	–	27.06	–	16.40	–	12.29	–	9.34
Standard Error	*1.16*		*0.99*		*0.77*		*0.87*		*0.85*
Pred. – True	**0.06**		**-0.20**		**-0.71**		**-0.34**		**-0.60**
$\widehat{\Pr}$(Att3=2)	29.50	–	33.53	–	28.19	–	22.35	–	19.94
Standard Error	*0.94*		*0.96*		*0.89*		*0.94*		*1.00*
Pred. – True	**-0.05**		**0.55**		**-0.81**		**0.84**		**-0.09**
$\widehat{\Pr}$(Att3=3)	24.11	–	39.41	–	55.41	–	65.36	–	70.72
Standard Error	*1.00*		*1.07*		*1.04*		*1.17*		*1.23*
Pred. – True	**-0.01**		**-0.36**		**1.53**		**-0.50**		**0.69**
$\widehat{\Pr}$(Att4=1)	33.52	–	28.75	–	23.86	–	20.83	–	15.30
Standard Error	*1.76*		*1.84*		*1.52*		*1.43*		*1.46*
Pred. – True	**1.03**		**-0.12**		**1.83**		**-0.20**		**0.94**
$\widehat{\Pr}$(Att4=2)	31.23	–	33.27	–	31.30	–	30.13	–	28.57
Standard Error	*4.57*		*4.80*		*4.78*		*4.87*		*4.94*
Pred. – True	**2.57**		**3.57**		**1.32**		**4.18**		**2.95**
$\widehat{\Pr}$(Att4=3)	35.25	–	37.97	–	44.84	–	49.03	–	56.13
Standard Error	*4.98*		*5.28*		*5.25*		*5.25*		*5.15*
Pred. – True	**-3.60**		**-3.46**		**-3.16**		**-3.99**		**-3.89**
$\widehat{\Pr}$(Attrit)	3.50	6.24	8.49	9.87	9.61	7.73	6.10	4.46	–
Standard Error	*0.37*	*0.38*	*0.50*	*0.45*	*0.54*	*0.55*	*0.42*	*0.56*	
Pred. – True	**-0.37**	**1.01**	**1.94**	**-3.98**	**2.81**	**1.55**	**-0.84**	**0.51**	

Attitudinal responses are jointly estimated as outcomes in **Model C**.

Attitudinal responses are only available every other year. Sample Size: 15,621.

Table A.4:
Model I, **Model E**, and **Model C** Estimated Response Parameters, Dependent Variable: Smoking

	Model I				**Model E**				**Model C**			
Variable	Pt Est	Std Err	P-value		Pt Est	Std Err	P-value		Pt Est	Std Err	P-value	
Constant	-17.77	15.76	0.26		-27.35	14.26	0.06		-31.38	10.08	0.00	*
Lagged Smoking	-1.51	1.85	0.41		-1.29	1.85	0.49		0.63	1.11	0.57	
Smoking Duration	3.40	1.61	0.04	*	3.30	1.65	0.05	*	1.60	0.90	0.08	
Smoking Experience	-0.34	0.70	0.63		0.09	0.69	0.89		0.45	0.58	0.43	
Lag. Smok. × Cig. Price	1.42	0.86	0.10		1.41	0.86	0.10		0.79	0.54	0.14	
Smok. Dur. × Cig. Price	-1.42	0.73	0.05	*	-1.49	0.74	0.04	*	-0.87	0.43	0.04	*
Smok. Exp. × Cig. Price	0.15	0.33	0.65		0.17	0.33	0.60		0.16	0.28	0.57	
Att1					0.13	0.07	0.06					
Att2					0.31	0.07	0.00	*				
Att3					0.18	0.07	0.01	*				
Att4					0.11	0.06	0.10					
Att *missing*					1.25	0.26	0.00	*				

Relative to alternative "No Smoking." Sample Size: 15,621.

Continued on the next page....

Table A.4 Continued:

Model I, **Model E**, and **Model C** Estimated Response Parameters, Dependent Variable: Smoking

	Model I				Model E				Model C			
Variable	Pt Est	Std Err	P-value		Pt Est	Std Err	P-value		Pt Est	Std Err	P-value	
Lives in Cincinatti	0.84	0.22	0.00	*	0.53	0.17	0.00	*	0.70	0.15	0.00	*
Lives in D.C.	-0.25	0.17	0.14		-0.22	0.13	0.08		-0.16	0.10	0.11	
Cigarette Price	0.38	0.96	0.69		0.05	0.94	0.96		0.78	0.65	0.22	
Race (0 white, 1 black)	-2.33	0.18	0.00	*	-1.78	0.12	0.00	*	-1.30	0.08	0.00	*
Number of Siblings	0.05	0.05	0.27		0.04	0.04	0.27		0.02	0.03	0.43	
Child Health Problem	0.15	0.18	0.40		0.06	0.14	0.67		0.01	0.11	0.95	
Year of the Study (1 to 9)	-1.43	0.68	0.04	*	-1.34	0.60	0.03	*	-1.36	0.51	0.01	*
$Year^2$ / 10	2.06	1.06	0.05		1.87	0.96	0.05		1.82	0.80	0.02	*
$Year^3$ / 100	-0.92	0.53	0.09		-0.84	0.49	0.09		-0.78	0.41	0.06	
Age	2.81	3.14	0.37		3.81	2.86	0.18		4.31	2.02	0.03	*
Age^2 / 10	-1.16	2.03	0.57		-1.89	1.86	0.31		-2.17	1.31	0.10	
Age^3 / 100	0.17	0.43	0.69		0.33	0.40	0.41		0.36	0.28	0.20	
Tall for Her Age	0.28	0.14	0.04	*	0.17	0.12	0.15		0.11	0.10	0.27	
Short for Her Age	-0.08	0.17	0.63		-0.07	0.13	0.60		-0.05	0.12	0.65	

Relative to alternative “No Smoking.” Sample Size: 15,621.

Continued on the next page....

Table A.4 Continued:

Model I, **Model E**, and **Model C** Estimated Response Parameters, Dependent Variable: Smoking

	Model I				Model E				Model C			
Variable	Pt Est	Std Err	P-value		Pt Est	Std Err	P-value		Pt Est	Std Err	P-value	
Number of Parents in HH	-0.21	0.15	0.17		-0.14	0.12	0.25		-0.14	0.10	0.14	
Parents' Education *bins*	-0.12	0.09	0.18		-0.12	0.07	0.08		-0.07	0.05	0.19	
Parents' Income *bins*	-0.19	0.07	0.01	*	-0.16	0.06	0.01	*	-0.11	0.05	0.01	*
Lives with Biological Mother	0.46	0.25	0.06		0.43	0.20	0.03	*	0.52	0.16	0.00	*
Lives with Biological Father	0.49	0.18	0.01	*	0.35	0.15	0.02	*	0.17	0.11	0.14	
Par. Smoking Habits *Yr 1*	0.28	0.12	0.03	*	0.29	0.10	0.00	*	0.20	0.08	0.01	*
Par. Drinking Habits *Yr 1*	0.16	0.10	0.09		0.11	0.07	0.12		0.08	0.05	0.13	
(Par. Age − Child Age)	-0.03	0.01	0.02	*	-0.02	0.01	0.14		-0.01	0.01	0.32	
Parent's BMI	0.03	0.01	0.01	*	0.02	0.01	0.03	*	0.01	0.01	0.06	
Child Hlth Prob. *missing*	-0.07	0.23	0.75		-0.11	0.19	0.57		0.08	0.15	0.60	
Parents' Income *missing*	-0.26	0.31	0.39		-0.30	0.26	0.24		-0.25	0.20	0.22	
Par. Smok. Hab. *missing*	0.72	0.26	0.01	*	0.52	0.22	0.02	*	0.27	0.17	0.11	
Parent's Age *missing*	-1.10	0.47	0.02	*	-0.56	0.34	0.10		-0.42	0.26	0.11	
Parent's BMI *missing*	1.01	0.37	0.01	*	0.70	0.29	0.01	*	0.43	0.22	0.05	

Relative to alternative "No Smoking." Sample Size: 15,621.

Table A.5: **Model C** Estimated Parameters for Outcomes Att1 and Att2

	Att1 = 2				**Att1 = 3**				**Att2 = 2**			
Variable	Pt Est	Std Err	P-value		Pt Est	Std Err	P-value		Pt Est	Std Err	P-value	
Constant	1.61	7.08	0.82		6.07	11.01	0.58		-6.59	8.52	0.44	
Lagged Smoking	3.03	1.40	0.03	*	3.71	2.12	0.08		2.16	1.65	0.19	
Smoking Duration	-1.52	0.96	0.12		-3.29	1.48	0.03	*	-2.63	1.17	0.02	*
Smoking Experience	0.84	0.77	0.27		2.37	1.14	0.04	*	1.06	0.90	0.24	
Lag. Smok. × Cig. Price	-1.28	0.68	0.06		-1.41	1.03	0.17		-0.92	0.80	0.25	
Smok. Dur. × Cig. Price	0.70	0.46	0.13		1.54	0.72	0.03	*	1.29	0.56	0.02	*
Smok. Exp. × Cig. Price	-0.36	0.37	0.32		-1.12	0.56	0.05	*	-0.55	0.43	0.21	
Lives in Cincinatti	0.32	0.11	0.00	*	0.48	0.18	0.01	*	0.27	0.13	0.03	*
Lives in D.C.	0.18	0.08	0.03	*	0.01	0.14	0.94		-0.05	0.11	0.62	
Cigarette Price	1.47	0.72	0.04	*	1.75	1.11	0.12		0.77	0.85	0.36	
Race (0 white, 1 black)	0.08	0.07	0.25		0.31	0.13	0.01	*	0.07	0.09	0.39	
Number of Siblings	-0.02	0.03	0.55		-0.11	0.04	0.02	*	-0.03	0.03	0.39	
Child Health Problem	-0.07	0.10	0.45		-0.15	0.17	0.36		0.01	0.13	0.97	
Year of the Study (1 to 9)	0.65	0.30	0.03	*	0.03	0.47	0.95		-0.31	0.35	0.38	
Year^2 / 10	-1.28	0.56	0.02	*	-0.59	0.88	0.50		0.58	0.66	0.38	
Year^3 / 100	0.62	0.31	0.05	*	0.37	0.50	0.45		-0.34	0.37	0.36	

Relative to alternative "Att1 = 1" or "Att2 = 1." Sample Size: 15,621.

Continued on the next page....

Table A.5 Continued: **Model C** Estimated Parameters for Outcomes Att1 and Att2

	Att1 = 2				**Att1 = 3**				**Att2 = 2**			
Variable	Pt Est	Std Err	P-value		Pt Est	Std Err	P-value		Pt Est	Std Err	P-value	
Age	-0.95	1.48	0.52		-2.40	2.32	0.30		1.09	1.78	0.54	
Age^2 / 10	0.62	1.01	0.54		1.77	1.60	0.27		-0.77	1.22	0.53	
Age^3 / 100	-0.12	0.23	0.60		-0.40	0.36	0.26		0.18	0.27	0.51	
Tall for Her Age	0.04	0.08	0.60		-0.22	0.15	0.14		-0.06	0.10	0.58	
Short for Her Age	0.22	0.09	0.02	*	-0.05	0.16	0.75		0.16	0.11	0.16	
Number of Parents in HH	0.06	0.08	0.46		0.05	0.14	0.70		0.04	0.11	0.73	
Parents' Education *bins*	0.05	0.05	0.34		0.02	0.08	0.76		0.02	0.06	0.68	
Parents' Income *bins*	-0.01	0.04	0.74		-0.15	0.07	0.03	*	-0.11	0.05	0.03	*
Lives with Biological Mother	0.00	0.15	0.98		-0.14	0.25	0.59		0.41	0.19	0.03	*
Lives with Biological Father	-0.09	0.10	0.38		0.09	0.17	0.60		-0.07	0.13	0.57	
Par. Smoking Habits *Yr 1*	0.19	0.07	0.01	*	0.24	0.12	0.04	*	0.15	0.09	0.09	
Par. Drinking Habits *Yr 1*	-0.03	0.05	0.52		-0.01	0.08	0.95		0.01	0.06	0.86	
(Par. Age − Child Age)	0.00	0.01	0.81		0.01	0.01	0.46		0.01	0.01	0.42	
Parent's BMI	0.01	0.01	0.01	*	0.03	0.01	0.00	*	0.01	0.01	0.04	*
Child Hlth Prob. *missing*	0.19	0.13	0.14		0.39	0.22	0.08		0.35	0.16	0.03	*
Parents' Income *missing*	-0.27	0.18	0.13		-0.83	0.30	0.01	*	-0.30	0.22	0.16	
Par. Smok. Hab. *missing*	-0.14	0.15	0.36		-0.11	0.24	0.63		-0.36	0.19	0.06	
Parent's Age *missing*	-0.19	0.23	0.41		-0.17	0.39	0.66		0.03	0.29	0.92	
Parent's BMI *missing*	0.42	0.20	0.03	*	0.98	0.33	0.00	*	0.36	0.25	0.16	

Relative to alternative "Att1 = 1" or "Att2 = 1." Sample Size: 15,621.

Table A.6: **Model C** Estimated Parameters for Outcomes Att2 and Att3

	Att2 = 3				**Att3 = 2**				**Att3 = 3**			
Variable	Pt Est	Std Err	P-value		Pt Est	Std Err	P-value		Pt Est	Std Err	P-value	
Constant	4.43	10.77	0.68		-16.24	7.87	0.04	*	-8.51	8.26	0.30	
Lagged Smoking	1.31	2.03	0.52		2.68	2.13	0.21		3.84	2.09	0.07	
Smoking Duration	-3.26	1.46	0.03	*	-3.85	1.60	0.02	*	-3.80	1.64	0.02	*
Smoking Experience	2.67	1.16	0.02	*	2.30	1.29	0.08		2.13	1.33	0.11	
Lag. Smok. × Cig. Price	-0.46	0.99	0.64		-1.35	1.04	0.20		-1.66	1.02	0.10	
Smok. Dur. × Cig. Price	1.62	0.71	0.02	*	1.85	0.76	0.01	*	1.80	0.78	0.02	*
Smok. Exp. × Cig. Price	-1.31	0.57	0.02	*	-1.06	0.60	0.08		-1.00	0.62	0.11	
Lives in Cincinatti	0.22	0.16	0.18		0.36	0.11	0.00	*	0.36	0.13	0.00	*
Lives in D.C.	-0.59	0.14	0.00	*	0.07	0.09	0.47		-0.05	0.11	0.64	
Cigarette Price	0.54	1.07	0.61		1.60	1.09	0.14		2.07	1.07	0.05	
Race (0 white, 1 black)	0.13	0.12	0.27		0.19	0.08	0.02	*	0.23	0.09	0.01	*
Number of Siblings	-0.05	0.04	0.27		0.00	0.03	0.95		0.00	0.03	0.99	
Child Health Problem	0.05	0.16	0.77		0.09	0.11	0.40		0.08	0.13	0.52	
Year of the Study (1 to 9)	-0.87	0.45	0.05		0.27	0.33	0.42		0.23	0.33	0.48	
$Year^2$ / 10	1.60	0.85	0.06		-0.44	0.66	0.51		-0.34	0.67	0.61	
$Year^3$ / 100	-0.96	0.48	0.04	*	0.20	0.39	0.60		0.21	0.39	0.59	

Relative to alternative "Att2 = 1" or "Att3 = 1." Sample Size: 15,621.

Continued on the next page....

Table A.6 Continued: **Model C** Estimated Parameters for Outcomes Att2 and Att3

	Att2 = 3				**Att3 = 2**			**Att3 = 3**			
Variable	Pt Est	Std Err	P-value		Pt Est	Std Err	P-value	Pt Est	Std Err	P-value	
Age	-0.95	2.26	0.67		2.42	1.68	0.15	0.07	1.75	0.97	
Age^2 / 10	0.45	1.57	0.77		-1.39	1.19	0.24	0.47	1.23	0.70	
Age^3 / 100	-0.06	0.35	0.87		0.28	0.27	0.31	-0.17	0.28	0.53	
Tall for Her Age	0.07	0.13	0.61		0.03	0.10	0.75	0.18	0.10	0.08	
Short for Her Age	0.07	0.15	0.63		-0.08	0.10	0.44	-0.10	0.12	0.40	
Number of Parents in HH	0.09	0.14	0.50		0.00	0.09	0.96	-0.16	0.11	0.14	
Parents' Education *bins*	-0.02	0.08	0.83		0.09	0.05	0.08	0.12	0.06	0.05	
Parents' Income *bins*	-0.16	0.07	0.02	*	0.03	0.04	0.47	0.09	0.05	0.09	
Lives with Biological Mother	0.45	0.25	0.07		0.14	0.17	0.40	0.09	0.20	0.67	
Lives with Biological Father	0.12	0.16	0.48		-0.13	0.11	0.24	0.10	0.13	0.46	
Par. Smoking Habits *Yr 1*	0.14	0.12	0.22		0.02	0.08	0.84	-0.03	0.09	0.74	
Par. Drinking Habits *Yr 1*	0.01	0.08	0.92		0.05	0.05	0.39	0.01	0.06	0.83	
(Par. Age − Child Age)	0.01	0.01	0.40		0.00	0.01	0.85	-0.02	0.01	0.03	*
Parent's BMI	0.03	0.01	0.00	*	0.00	0.01	0.50	-0.01	0.01	0.41	
Child Hlth Prob. *missing*	0.30	0.22	0.17		0.07	0.14	0.63	0.18	0.17	0.29	
Parents' Income *missing*	-0.56	0.29	0.05		-0.04	0.19	0.85	0.13	0.23	0.56	
Par. Smok. Hab. *missing*	-0.23	0.25	0.37		0.05	0.17	0.77	-0.04	0.19	0.85	
Parent's Age *missing*	-0.34	0.39	0.38		-0.10	0.26	0.71	-0.83	0.31	0.01	*
Parent's BMI *missing*	1.06	0.34	0.00	*	-0.15	0.22	0.49	0.07	0.26	0.77	

Relative to alternative "Att2 = 1" or "Att3 = 1." Sample Size: 15,621.

Table A.7: **Model C** Estimated Parameters for Outcome Att4 and Attrition

	Att4 = 2				**Att4 = 3**				**Attrition**			
Variable	Pt Est	Std Err	P-value		Pt Est	Std Err	P-value		Pt Est	Std Err	P-value	
Current Period Smoking									0.31	0.13	0.02	*
Constant	-6.72	7.80	0.39		33.75	9.40	0.00	*	-42.90	8.48	0.00	*
Lagged Smoking	0.81	1.73	0.64		-2.35	1.68	0.16		1.66	1.42	0.24	
Smoking Duration	-0.80	1.22	0.51		0.22	1.19	0.85		-0.15	1.24	0.91	
Smoking Experience	0.26	0.95	0.79		0.62	0.94	0.51		-0.03	0.95	0.98	
Lag. Smok. × Cig. Price	-0.46	0.84	0.59		1.17	0.81	0.15		-0.69	0.73	0.34	
Smok. Dur. × Cig. Price	0.40	0.58	0.50		-0.09	0.57	0.88		0.05	0.62	0.94	
Smok. Exp. × Cig. Price	-0.13	0.45	0.77		-0.34	0.45	0.45		0.02	0.48	0.97	
Lives in Cincinatti	0.10	0.11	0.38		0.49	0.13	0.00	*	0.36	0.14	0.01	*
Lives in D.C.	-0.03	0.09	0.77		0.25	0.11	0.02	*	0.17	0.10	0.10	
Cigarette Price	0.55	0.89	0.53		-0.54	0.86	0.53		1.03	0.80	0.20	
Race (0 white, 1 black)	-0.12	0.08	0.12		-0.36	0.09	0.00	*	-0.19	0.08	0.02	*
Number of Siblings	0.02	0.03	0.53		-0.06	0.04	0.11		0.02	0.03	0.43	
Child Health Problem	0.09	0.11	0.42		0.21	0.12	0.10		0.15	0.11	0.18	
Year of the Study (1 to 9)	0.08	0.32	0.82		0.81	0.38	0.03	*	0.19	0.47	0.69	
Year^2 / 10	-0.13	0.63	0.84		-1.91	0.72	0.01	*	-0.03	0.91	0.98	
Year^3 / 100	0.09	0.36	0.81		1.22	0.40	0.00	*	-0.54	0.56	0.34	

Relative to alternative "Att4 = 1" or "Does Not Attrit." Sample Size: 15,621.

Continued on the next page....

Table A.7 Continued: **Model C** Estimated Parameters for Outcome Att4 and Attrition

Age	1.37	1.64	0.40		-6.93	1.99	0.00	*	7.82	1.83	0.00 *
Age^2 / 10	-0.89	1.14	0.44		4.93	1.37	0.00	*	-5.37	1.29	0.00 *
Age^3 / 100	0.20	0.26	0.45		-1.11	0.31	0.00	*	1.26	0.30	0.00 *
Tall for Her Age	-0.06	0.09	0.50		0.04	0.10	0.69		0.07	0.10	0.51
Short for Her Age	0.05	0.11	0.61		0.22	0.11	0.05		0.03	0.11	0.76
Number of Parents in HH	0.02	0.09	0.84		-0.21	0.10	0.04	*	0.11	0.09	0.25
Parents' Education *bins*	-0.02	0.05	0.65		-0.09	0.06	0.14		-0.03	0.05	0.53
Parents' Income *bins*	-0.01	0.04	0.83		0.06	0.05	0.25		-0.15	0.04	0.00 *
Lives with Biological Mother	-0.02	0.17	0.91		0.03	0.19	0.87		0.26	0.14	0.08
Lives with Biological Father	-0.31	0.11	0.01	*	-0.14	0.13	0.29		-0.07	0.11	0.55
Par. Smoking Habits *Yr 1*	-0.10	0.08	0.19		-0.07	0.09	0.41		0.00	0.08	0.98
Par. Drinking Habits *Yr 1*	-0.05	0.05	0.40		0.00	0.06	0.98		0.14	0.06	0.02 *
(Par. Age − Child Age)	-0.01	0.01	0.30		-0.02	0.01	0.02	*	-0.04	0.01	0.00 *
Parent's BMI	0.00	0.01	0.74		0.01	0.01	0.10		0.00	0.01	0.54
Child Hlth Prob. *missing*	0.15	0.14	0.31		0.14	0.17	0.40		-0.29	0.13	0.03 *
Parents' Income *missing*	0.08	0.19	0.68		0.54	0.22	0.01	*	-0.50	0.19	0.01 *
Par. Smok. Hab. *missing*	-0.37	0.16	0.02	*	-0.44	0.19	0.02	*	0.45	0.15	0.00 *
Parent's Age *missing*	-0.34	0.25	0.18		-0.60	0.30	0.05	*	-0.78	0.28	0.01 *
Parent's BMI *missing*	0.21	0.21	0.33		0.42	0.26	0.10		0.38	0.22	0.09

Relative to alternative "Att4 = 1" or "Does Not Attrit." Sample Size: 15,621.

Table A.8: **Model I** and **Model E** Estimated Attrition Parameters

	Model I: Attrition				Model E: Attrition			
Variable	Pt Est	Std Err	P-value		Pt Est	Std Err	P-value	
Current Period Smoking	-0.45	0.33	0.18		0.03	0.20	0.89	
Current Period Smoking	-0.45	0.33	0.18		0.03	0.20	0.89	
Constant	-62.47	122.20	0.61		-41.62	14.28	0.00	*
Lagged Smoking	1.17	1.55	0.45		1.40	1.44	0.33	
Smoking Duration	0.45	1.37	0.74		-0.01	1.27	0.99	
Smoking Experience	-0.43	1.04	0.67		-0.20	0.97	0.84	
Lag. Smok. × Cig. Price	-0.47	0.79	0.55		-0.65	0.74	0.38	
Smok. Dur. × Cig. Price	-0.18	0.69	0.80		0.03	0.63	0.96	
Smok. Exp. × Cig. Price	0.15	0.52	0.78		0.04	0.49	0.93	
Att1					0.10	0.07	0.15	
Att2					-0.03	0.07	0.66	
Att3					0.05	0.07	0.41	
Att4					-0.06	0.06	0.37	
Att *missing*					0.34	0.24	0.15	
Lives in Cincinatti	0.41	0.17	0.01	*	0.34	0.14	0.02	*
Lives in D.C.	0.18	0.15	0.22		0.14	0.11	0.18	
Cigarette Price	0.83	0.89	0.35		1.01	0.81	0.21	
Race (0 white, 1 black)	-0.45	0.13	0.00	*	-0.29	0.10	0.00	*
Number of Siblings	0.04	0.03	0.27		0.02	0.03	0.43	
Child Health Problem	0.29	0.14	0.04	*	0.15	0.12	0.18	
Year of the Study (1 to 9)	-0.73	1.08	0.50		-0.16	0.55	0.77	
Year^2 / 10	2.04	2.18	0.35		0.68	1.08	0.53	
Year^3 / 100	-2.03	1.43	0.15		-0.97	0.66	0.14	

Relative to alternative "Does Not Attrit." Sample Size: 15,621.

Continued on the next page....

Table A.8 Continued: **Model I** and **Model E** Estimated Attrition Parameters

	Model I: Attrition				Model E: Attrition			
Variable	Pt Est	Std Err	P-value		Pt Est	Std Err	P-value	
Age	18.49	9.27	0.05	*	7.73	3.04	0.01	*
Age^2 / 10	-12.62	6.28	0.04	*	-5.32	2.11	0.01	*
Age^3 / 100	2.92	1.41	0.04	*	1.26	0.48	0.01	*
Tall for Her Age	0.16	0.11	0.16		0.07	0.10	0.49	
Short for Her Age	0.03	0.12	0.84		0.03	0.11	0.80	
Number of Parents in HH	0.16	0.11	0.14		0.11	0.10	0.27	
Parents' Education *bins*	-0.07	0.07	0.27		-0.05	0.06	0.42	
Parents' Income *bins*	-0.22	0.05	0.00	*	-0.16	0.05	0.00	*
Lives with Biological Mother	0.29	0.17	0.09		0.22	0.15	0.14	
Lives with Biological Father	-0.04	0.14	0.78		-0.04	0.12	0.76	
Par. Smoking Habits *Yr 1*	-0.02	0.09	0.86		0.02	0.08	0.80	
Par. Drinking Habits *Yr 1*	0.18	0.08	0.01	*	0.14	0.06	0.02	*
(Par. Age − Child Age)	-0.07	0.02	0.00	*	-0.04	0.01	0.00	*
Parent's BMI	0.01	0.01	0.27		0.01	0.01	0.34	
Child Hlth Prob. *missing*	-0.37	0.16	0.02	*	-0.35	0.14	0.01	*
Parents' Income *missing*	-0.69	0.24	0.00	*	-0.52	0.20	0.01	*
Par. Smok. Hab. *missing*	0.62	0.19	0.00	*	0.51	0.16	0.00	*
Parent's Age *missing*	-1.27	0.41	0.00	*	-0.81	0.29	0.00	*
Parent's BMI *missing*	0.56	0.27	0.04	*	0.46	0.23	0.05	

Relative to alternative "Does Not Attrit." Sample Size: 15,621.

Bibliography

Arcidiacono, Peter, Holger Sieg, and Frank Sloan, 2007, Living rationally under the volcano? an empirical analysis of heavy drinking and smoking, *International Economic Review* 48, 37–65.

Boxall, Peter C., and Wiktor L. Adamowicz, 2002, Understanding heterogeneous preferences in random utility models: A latent class approach, *Environmental and Resource Economics* 23, 421–446.

Breffle, William, Edward Morey, and Jennifer Thacher, 2011, A joint latent-class model: Combining likert-scale preference statements with choice data to harvest preference heterogeneity, *Environmental and Resource Economics* 50, 83–110.

Cawley, John, Sara Markowitz, and John Tauras, 2004, Lighting up and slimming down: the effects of body weight and cigarette prices on adolescent smoking initiation, *Journal of Health Economics* 23, 293–311.

Chintagunta, Pradeep K, Dipak C Jain, and Naufel J Vilcassim, 1991, Investigating heterogeneity in brand preferences in logit models for panel data, *Journal of Marketing Research* pp. 417–428.

Conti, Gabriella, James J. Heckman, Hedibert F. Lopes, and Remi Piatek, 2010, Constructing economically justified aggregates: An application to the early origins of health, *(Working Paper)*.

DeCicca, Philip, Donald Kenkel, and Alan Mathios, 2002, Putting out the fires: Will higher taxes reduce the onset of youth smoking?, *Journal of Political Economy* 110, 144–169.

Fletcher, Jason M., Partha Deb, and Jody L. Sindelar, 2009, Tobacco use, taxation and self control in adolescence, *National Bureau of Economic Research Working Paper Series* No. 15130.

Gilleskie, Donna B., and Koleman S. Strumpf, 2005, The behavioral dynamics of youth smoking, *The Journal of Human Resources* 40, 822–866.

McCullagh, P., and J.A. Nelder, 1989, *Generalized Linear Models* (Chapman and Hall).

McFadden, Daniel, 1986, The choice theory approach to market research, *Marketing Science* 5, 275–297.

Morey, Edward, Jennifer Thacher, and William Breffle, 2006, Using angler characteristics and attitudinal data to identify environmental preference classes: A latent-class model, *Environmental and Resource Economics* 34, 91–115.

Mroz, Thomas A., 1999, Discrete factor approximations in simultaneous equation models: Estimating the impact of a dummy endogenous variable on a continuous outcome, *Journal of Econometrics* 92, 233–274.

Zellner, Arnold, 1962, An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias, *Journal of the American Statistical Association* 57, 348–368.